Multivariate Risikomanagement

ABSCHLUSSARBEIT

zur Erlangung des akademischen Grades

Master of Science

(M.Sc.)

im Masterstudiengang Statistik

an der Wirtschaftwissenschaftlichen Fakultät

Humboldt-Universität zu Berlin

von

Ying Chen

geboren am 08.12.1975 in Kaifeng, Volksrepublik China

Gutachter:

Prof. Dr. Wolfgang Härdle
Prof. Dr. Vladimir Spokoiny
eingereicht am 6th February 2005

Multivariate Risk Management

A Master Thesis Presented

by

Ying Chen

(169607)

to

Prof. Dr. Wolfgang Härdle

CASE - Center of Applied Statistics and Economics

Humboldt-Universität zu Berlin

in partial fulfillment of the requirements

for the degree of

Master of Science

Berlin, 6th February 2005

Declaration of Authorship

I hereby confirm that I have authored this master thesis independently and without use of others than the indicated resources. All passages, which are literally or in general matter taken out of publications or other resources, are marked as such.

Ying Chen
Berlin, 6th February 2005

ABSTRACT

In this thesis we propose a risk management methodology to high-dimensional financial portfolios. Instead of estimating the joint density of the portfolios in a high-dimensional space, we are encouraged by using the independent component analysis (ICA) to decompose the dependent risk factors to a linear transformation of independent components (ICs). The marginal density and the volatility process of each IC are estimated in a univariate dimension. Thereafter the joint densities and the dependence structures of the ICs and the original risk factors can be calculated using the statistical property of the independence and its linear transformation. We assume the marginal densities of ICs belong to the generalized hyperbolic (GH) distribution family since this family possesses semi-heavy tails and mimics the empirical distributions of the ICs appropriately. Further we implement a nonparametric adaptive methodology to estimate the local volatilities of ICs based on a homogeneity test. In order to check the reliability of the proposed methodology, we consider a portfolio in our study: a 2-dimensional exchange rates DEM/USD and GBP/USD with 4 different trading strategies. The empirical studies show that the performance of the VaR forecast using the proposed methodology is better than the popular Delta-Gamma-Normal model. All calculations and simulations are able to be recalculated with the software XploRe.

Keywords: independent component analysis, generalized hyperbolic distribution, adaptive volatility, Value-at-Risk

ACKNOWLEDGEMENT

Special thanks to Professor Dr. Wolfgang Härdle and Prof. Dr. Vladimir Spokoiny for their encouragements and suggestions on my study.

NOTATION

$\|\| \cdot \|\|$	norm of a vector
$\alpha, \beta, \delta, \mu, \lambda$	generalized hyperbolic distribution parameters
Σ_t	covariance of portfolio
σ_t	volatility of portfolio
ε_t	stochastic error term
ϱ	proportion of the nongaussianity explained by the chosen independent components
$\mathsf{E}[x]$	expected value of the random variable x
I	identity matrix
R^d	d-dimensional real number space
N	Gaussian distribution
$A^\top$	transpose of the vector of matrix A
a	VaR probability level
b	weighting vector of portfolio
$\text{cum}(x_i, x_j, x_k, x_l)$	4-th order cumulant of the random variable x
f_y	probability density of the random variable y
$H(x)$	entropy of the random variable x
$J(x)$	negentropy of the random variable x
$\text{kurt}(x)$	kurtosis of the random variable x
R_t	profit and loss of portfolio
$s(t)$	genuine independent time series
$x(t)$	time series observed in the market
$y(t)$	independent components

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

After the breakdown of the fixed exchange rate system of the Bretton Woods Agreement in 1971, a sudden increase of volatility was observed in the financial markets. The following boom of financial derivatives accelerated the turbulence of the markets. Besides, the incoming scale of losses astonished the world and pushed the development of sound risk management systems. One challenging task in risk management is to measure and manage risks properly. Financial risks have many sources and are typically mapped into a stochastic framework where various kinds of risk measures such as Value at Risk (VaR), expected shortfall (ES), lower partial moments are calculated. Among them, VaR has become the standard measure of the market risk since J.P. Morgan launched RiskMetrics in 1994, making the analysis of VaR simple, Jorion (2001). VaR is briefly to say the cutoff loss over a target horizon that will happen $100 \cdot a\%$ of the time. Mathematically VaR at a probability level a is defined as:

$$VaR_{a,t} = F_t^{-1}(a), \tag{1.1}$$

where F_t^{-1} is the inverse function of the cumulative distribution function (cdf) of the profit and loss (P&L) at time t, Franke, Härdle and Hafner (2004). VaR has a drawback that it is not sub-additive, i.e. the risk of an aggregated portfolio is not always smaller than the sum of the risks of the individual components. As a result it is not only inconsistent with diversification but also a wrong risk measure for allocating capital changes and measuring capital adequacy. Nevertheless, VaR is still considered as an industrial standard of risk measure and widely used in risk management. Since from the viewpoints of shareholder and management, the bankruptcy is independent of the size of loss that triggers the default, Taleb (2001). The importance of VaR was even reinforced after it was used by the central banks to govern and supervise the capital adequacy of the banks in the Group of Ten (G10) countries in 1995. Therefore in this thesis we mainly study the problems open in the VaR calculation given a high-dimensional portfolio. Note that the methodology proposed here is applicable to other risk measures as well.

The heteroscedastic model is applied in risk management to illustrate the stochastic property of the return of the portfolio:

$$R_t = b^\top x_t = b^\top \Sigma_t^{1/2} \varepsilon_t, \tag{1.2}$$

where R_t denotes the return of the portfolio at time point t, $x_t \in R^d$ the returns of the

individual financial instruments, $b = (b_1, \cdots, b_d)^\top$ the weighting vector of the portfolio, Σ_t the covariance of the portfolio and $\varepsilon_t \in R^d$ an independent and identical distributed (i.i.d.) stochastic error term. The terms x_t and b are generally given in advance while the covariance and the stochastic term need to be estimated. It is clear that the covariance estimation and the distribution assumption of the stochastic process are of the greatest importance for the success of the VaR calculation. In a low dimensional space, one can use non- and semiparametric methods to estimate the joint density and the covariance. But it becomes immediately infeasible if we consider a high-dimensional portfolio due to the curse of dimensionality problem. On the other hand, portfolios holding by financial institutions often consist of hundreds or even thousands of financial instruments. Here we propose an easy way to solve the high-dimensional problem in a univariate space. Then the two estimations of the VaR calculation, i.e. the joint density and the covariance estimation can be done fast and easily. Before introducing our idea, it is convenient to review three models widely used to calculate the portfolio VaR in the current

Historical simulation model applies the current weights of the portfolio to the historical returns and constructs a hypothetical portfolio process. This model actually reduces the dimension of the portfolio from d to 1 and hence avoids the high-dimensional estimations. The success of this model however requires a large amount of historical data, and further a new hypothetical series needs to be constructed again if the weights of the portfolio are changed. In this sense, it is weakened by the information loss of the risk contributions of the individual instruments and the repetition of the density and the volatility estimations. Monte Carlo (MC) simulation model and Delta-Gamma-Normal model, on the contrary, deal with the multivariate data. The former simulates the stochastic processes of all variables of interest and estimates the joint density and the covariance based on the simulated series. This model gives the most precise VaR calculation but its computational time is very expensive. The Delta-Gamma-Normal model speeds up the calculation by assuming that the involved risk factors are normally distributed. It is argued that the central limit theorem (CLT) supports the normality assumption theoretically, i.e. the cumulative distribution function of the sum of the i.i.d. variables approaches a normal distribution. However the risk factors have a dependent feature which violates at least the independent condition of the CLT. Further a question arises in the normality framework: How close is the unknown portfolio distribution to the normal one, especially in the tails? Empirically, the normal model losses accuracy by underestimating the risk level.

Here we explore a different train of thought to calculate the VaR of a high-dimensional portfolio: Use the desirable property of independence and apply independent component analysis (ICA) to the portfolio. The statistical concept - independence- is defined in terms of probability densities.

DEFINITION 1 (Independence):

Random variables $y_1, \cdots, y_d$ are independent if and only if:

$$f_{y_1,\cdots,y_d} = \prod_{i=1}^{d} f_{y_i} \tag{1.3}$$

where $f_{y_1,\cdots,y_d}$ denotes the joint density and f_{y_i} the marginal density.

Independent variables have a lot of desirable properties. Above all, their joint density as given in the definition is the product of their marginal densities. The marginal densities of the independent variables can be estimated using statistical methods such as the nonparametric kernel density estimation. The covariance of the independent variables is a diagonal matrix.

$$\Sigma_{t,y} \quad = \quad \text{diag}(\sigma^2_{t,y_1}, \cdots, \sigma^2_{t,y_d})$$

These properties impelled us to apply ICA to find a matrix W which transforms the original dependent vector $x(t) = \{x_1(t), \cdots, x_d(t)\}^\top$, the P&L of the portfolio, to an independent vector $y(t) = \{y_1(t), \cdots, y_d(t)\}^\top$. It can be formulated as:

$$y(t) = Wx(t). \tag{1.4}$$

Given the value of the $d \times d$ dimensional matrix W and assumed W is a nonsingular matrix, i.e. it has an inverse matrix $W^{-1} = A$, the joint density of the dependent vector $x(t)$, according to the Jacobian transformation (Härdle and Simar, 2003), is

$$f_x \quad = \quad \text{abs}(|W|) f_y(Wx)$$

The covariance of the dependent vector is a linear transformation of the diagonal matrix:

$$\Sigma_{t,x} \quad = \quad A\Sigma_{t,y}A^\top.$$

The ICA idea can be illustrated by a simple example. We independently simulate one sinus function $sin(t)$ with $t = 1, \cdots, 500$, one uniform random variable and one generalized hyperbolic random variable, each with 500 observations. The time series of the independent vector is displayed in the top row of Figure 1.1. Given a time-invariant matrix $A = \begin{pmatrix} 2 & -3 & 0.5 \\ -1 & -8 & 2 \\ 12 & -5 & 1 \end{pmatrix}$, we obtain a dependent vector $x(t)$, the product of the matrix A and the independent vector. The time series of $x(t)$ is shown in the middle row of Figure 1.1. These plots can be imagined as the returns of the portfolio observed in the market. They are dependent and display a similar patten in general. But in fact, they are con-

trolled by 3 hidden independent factors. We use an ICA algorithm proposed by Cardoso and Souloumiac (1993), joint approximate diagonalization of eigenmatrices (JADE), to estimate the mixing matrix A and the ICs. The time series of the estimates of ICs is given in the bottom of Figure 1.1. This example provides an evidence that ICs can be found using ICA if they do exist and the dependent vector can be explained by a linear transformation of them. In financial markets, the movements of financial returns are often ignited by the common economic factors such as interest rate. Therefore, ICA is economic realistic. On the condition that ICs are obtained, we can calculate the high-dimensional VaR in a univariate space. Remember there are still two tasks in a line: distribution assumption and volatility estimation.

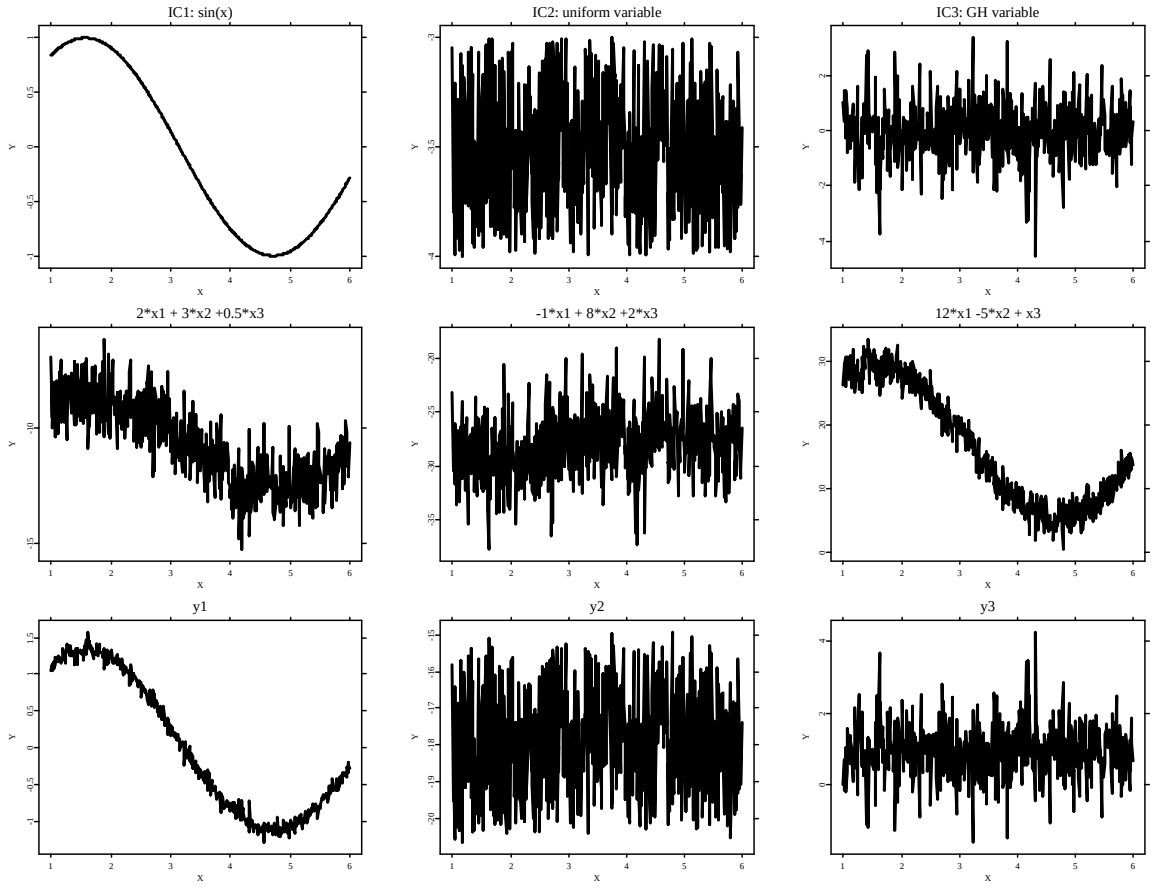

Fig. 1.1: Simulated independent components: $sin(t)$, uniform variable and GH variable (above). The dependent vector after a linear transformation (middle). The independent components estimated by applying JADE algorithm (bottom).

ICAsim.xpl

1. In risk management, the tails of the density are of vital importance. Empirical stud-

ies have shown that most financial risk factors have a leptokurtic distribution which includes a high peak and fat tails. Normality is obviously not a suitable assumption in this case. The nonparametrically estimated density of stock return, for example, obviously deviates from the normal density. Although the nonparametric density estimation gives the most accurate fit in average, it has a poor performance in the tails where few observations are available. Another drawback of the nonparametric estimation comes from the fact that the quantile or the inverse of the density without a form is difficult to compute. On the contrary, parametric estimation gives the density form explicitly and is therefore preferable, if it captures the heavy tail property. There has been keen competition among various heavy-tailed distribution families. For example, conditional Gaussian distribution was used in Jaschke and Jiang (2002), which is able to mimic the fat tails well at a moderate VaR (e.g. 95%) confidence level. However it underestimates the risk of an extreme event at 99% confidence level. In the recent, hyperbolic distributions have been attracting the attentions of researchers. Eberlein, Kallsen and Kristen (2003), Chen, Härdle and Jeong (2005) calculated VaR at different confidence levels assuming the risk factor follows generalized hyperbolic (GH) distribution. In both studies, the GH density gave more accurate VaR values than the normal one. Without doubt, there are other heavy-tailed distributions which can be used to mimic the empirical density. For example, Laplace distribution, Cauchy distribution and t-distribution. Compared to them, the GH distribution has a desirable tail behavior and this helps us to estimate the volatility of IC adaptively. Thus we assume in our study that ICs belong to the GH distribution family. We consider the daily foreign exchange (FX) rates DEM/USD and GBP/USD from 1979-12-01 to 1994-04-01. Figure 1.2 compares the nonparametric density, the normal density and the hyperbolic (HYP) density estimations of the first IC of the FX portfolio. It shows that the HYP fit almost coincides the nonparametric density.

2. Last but not least, how can we estimate the covariance in risk management? If the variables are independent, the covariance is a diagonal matrix. Thus we can estimate these elements in the diagonal line, i.e. the volatilities of ICs separately. The development of volatility estimation can be summarized as:

$$\sigma \quad \rightarrow \quad \sigma_t = m(\cdot) \quad \rightarrow \quad \sigma_t = m_t(\cdot),$$

Briefly to say, volatility was firstly assumed to be constant, e.g. in the Black Scholes (BS) model introduced in 1973. However it is rejected by the "smile" of the implied volatility and the volatility clustering displayed in the markets. Thereafter volatility is regarded to be time dependent. Firstly one used a time-invariant function m to estimate the volatility such as ARCH (Engle, 1995), GARCH (Bollerslev, 1995) and stochastic volatility models (Harvey, Ruiz and Shephard, 1995). Such a model has an inherent weakness: For a long time series the form of the volatility model may be vari-

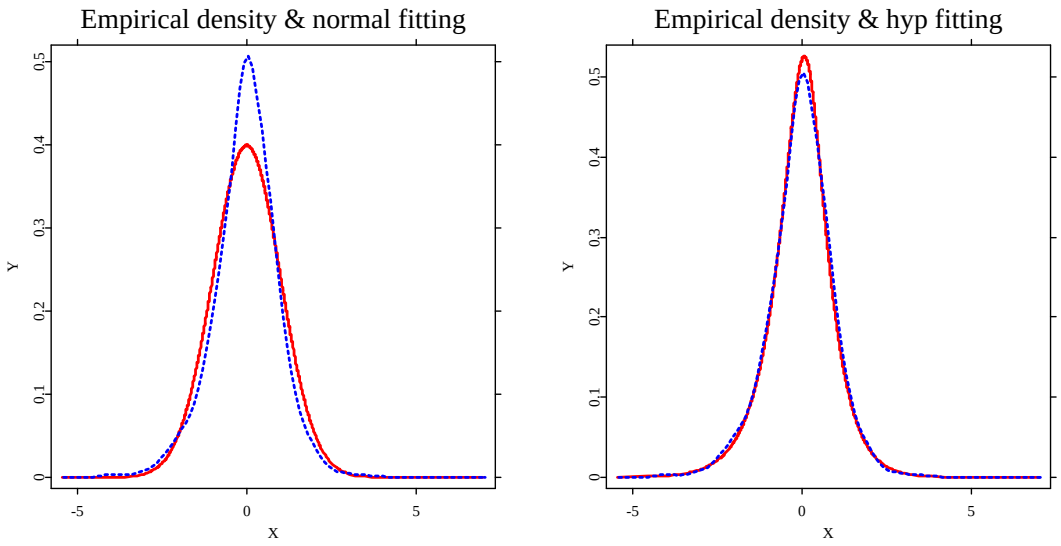

Fig. 1.2: Graphical comparison of the density estimations of one independent component on the basis of a foreign exchange portfolio (DEM/USD & GBP/USD)$^\top$ from 1979/12/01 to 1994/04/01 (3719 observations). The dotted line is the kernel density estimate. The normal fit is shown on the left panel whereas the hyperbolic fit on the right.

ICAfx.xpl

ant with a very high possibility as the volatility itself. In order to avoid this potential miss-specification problem, it is plausible to use a more flexible tool by providing a data-driven "local" model, i.e. $\sigma_t = m_t(\cdot)$ is used. Mercurio and Spokoiny (2003) introduced a model to estimate the volatility adaptively. The algorithm assumes that the volatility is constant in a short time interval although it can be volatile in a long run. The theory of their proposed "local constant" model was however based on the normality assumption of the risk factors. Chen et al. (2005) extended the algorithm based on the GH random variable. In this thesis, we will follow their study to estimate the volatility.

In a summary, this master thesis is motivated by the above three research clues. We implement ICA to a high-dimensional portfolio. Based on the ICs, we estimate the volatility and the GH distribution parameters of each series. The VaRs of the portfolio return at different levels are then estimated by Monte Carlo simulations.

Chapter 2 gives an overview on ICA. The properties of ICA and various algorithms are discussed. Among them, two algorithms are intensively compared. For example, the JADE algorithm used in the above example was very popular after its introduce. But it is only computationally fast for a low-dimensional data. On the contrary, FastICA is fast in the high-dimensional study.

Chapter 3 describes the properties of the univariate GH distribution and its subclasses: Hyperbolic (HYP) distribution and normal inverse Gaussian (NIG) distribution. These two subclasses are suitable to explain the empirical marginal distributions of the financial ICs. The GH distribution parameters are estimated using the maximum likelihood (ML) and numerical optimization methods. Furthermore, the tail behavior of the several heavy-tailed distributions will be compared. One can see that GH distribution has a lighter tail than others and this desirable property will help us to estimate volatility adaptively.

Chapter 4 gives details on the adaptive volatility estimation based on the GH distribution. A time homogeneous test is derived to specify the short homogeneous interval where the volatility is almost constant. The "local" constant volatility is then estimated by the average value of the squared returns in this interval.

The validation of the proposed VaR methodology will be discussed in Chapter 5 based on one portfolio family: a 2-dimensional exchange rates DEM/USD and GBP/USD with 4 different trading strategies.

Finally we will conclude in Chapter 6. All the pictures may be recalculated and redrawn using the indicated links to an XploRe Quantlet Server. The codes and data used in our study are available at http://ise.wiwi.hu-berlin.de/∼ychen/ICA/.

2. INDEPENDENT COMPONENT ANALYSIS

The study of independent component analysis (ICA) started from the early 1980s, when this method was first used in acoustics study for the deconvolution of signals. The well-known example is the "cocktail party problem". In a noisy party, researchers use ICA to distinguish voices of different people from several mixed acoustic waves recorded. These voices are naturally considered as independent sources. Promoted by the success in acoustics, ICA has been applied in different areas such as the brain imaging study Duann, Jung, Kuo, Yeh, Makeig, Hsieh and Sejnowski (2002) and the telecommunication study Ristaniemi, Raju and Karhunen (2002). The implementation of ICA in financial time series is comparably late. The very earliest ICA study in finance, as we know, can be traced back to Back and Weigend (1998). They considered the daily returns of the 28 largest Japanese stocks from 1986 to 1989 and compared the explanation abilities of the overall stock price using the independent components (ICs) and the principal components (PCs) respectively. It is found that ICs give more precise approximation than the PCs. There are however few studies of ICA in risk management. We are inspired by the outstanding achievements of ICA and intend to apply the method in the risk management of financial portfolio. For a complete and deep introduction of ICA, we refer to Hyvärinen, Karhunen and Oja (2001).

In all the financial related fields, it is very important to estimate the joint distribution of financial instruments appropriately, since traders and managers make decisions based on the distribution of their portfolios. There are a lot of contributions on the univariate density estimation. One can for example choose a heavy-tailed distribution which can mimic the leptokurtic feature of the financial time series or more generally, estimate the density nonparametrically. However it becomes more complicated if one considers a high-dimensional portfolio. Many estimation methods are although successful in a univariate case, they are very often error-prone or even infeasible to estimate the joint density in a high-dimensional space. This may result from the inborn features of these estimation methods. For example the nonparametric kernel density estimation has poor performance due to the curse dimensionality problem. Therefore, two solutions are suggested: dimension reduction and normality assumption. For the basic concepts of multivariate statistical analysis, we refer to Härdle and Simar (2003).

Given a d-dimensional data with d very large, the intuitive thinking is to reduce the dimensionality using statistical methods. PCA is e.g. widely used, choosing the first m

PCs based on their explanation degrees of the covariance. In this case, we can estimate the statistical properties such as the joint density of the high-dimensional data in a low m-dimensional space. There are however several problems when we apply PCA in the risk management. Above all, risk measures concern the tail of the joint density which is normally measured by the 4-th moment such as kurtosis. The dimension reduction based on the covariance may lose important information in the tail. Secondly, if the value of m is still larger than e.g. 3, the curse dimensionality problem appears again. Thus some assumptions are proposed before the dimension reduction method applied, e.g. the normality of the data. The central limit theorem (CLT) shows that the distribution of a sum of independent and identical distributed (i.i.d.) random variables converges to a Gaussian distribution. Briefly to say, if investor holds enough financial instruments in his portfolio and these instruments are i.i.d, the portfolio is normally distributed. If the high-dimensional portfolio is approximately normally distributed, then its first two moments, i.e. mean and covariance can control the joint distribution completely. Moreover, the decomposition of PCA based on the covariance gives an easy way to estimate the distribution of the portfolio. Since the decomposed uncorrelated normal variables are independent, the joint density of the data is the product of the marginal normal densities and the covariance is a diagonal matrix. Unfortunately, the normality assumption is unrealistic since financial instruments are highly correlated and leptokurtic distributed. Therefore, the risk management model based on the normality assumption often underestimates the risk level. But the essence of the normality assumption, i.e. using the statistical properties of independence, encourages us to apply ICA here.

The pith of ICA is to find ICs based on the observed vector $x(t)$. The ICs can be formulated as a linear function of $x(t)$. Given an independent portfolio $y(t) = \{y_1(t), \cdots, y_d(t)\}^\top$ and assumed that $y(t) = Wx(t)$, where W is a $d \times d$ nonsingular matrix. We can firstly estimate the marginal densities and the volatilities of the elements y_i separately for $i = 1, \cdots, d$. Since $y(t)$ is independent from each other, we obtain:

1. The joint density of $y(t)$ is the product of the marginal densities f_{y_i}. And further the density of $x(t)$ can be obtained according to the Jacobian transformation:

$$\begin{aligned} f_y &= \prod_{i=1}^{d} f_{y_i} \\ f_x &= \text{abs}(|W|) f_y(Wx) \end{aligned}$$

2. The covariance of the portfolio, which describes the dependence structure of the portfolio, is a diagonal matrix with the volatilities σ^2_{t,y_i} as the diagonal elements. The covariance of $x(t)$ is a linear transformation of the diagonal matrix:

$$\Sigma_{t,y} = \text{diag}(\sigma^2_{t,y_1}, \cdots, \sigma^2_{t,y_d})$$

$$\Sigma_{t,x} \quad = \quad W^{-1}\Sigma_{t,y}W^{-1^\top}$$

Briefly to say, if the elements of the portfolio are independent, the high-dimensional estimations can be simply calculated by these univariate estimations. The basic ICA assumes that the linear function is perfectly fulfilled and the ICs have the same dimension as the observed vector. These assumptions sound too optimistic. In order to avoid a possible miss-specification of the basic ICA, more general ICA models were introduced, e.g. the reduced ICA and the noisy ICA. The former involves a dimension reduction of the ICs, i.e. $y(t) \in R^m$ with $m \leq d$. This model is useful if one has the knowledge of the ICs' dimension in advance, for example it is practical in the brain imagining study. Very often, PCA is used in the reduced ICA model to specify the dimension based on the covariance of the portfolio. In our study, it is firstly hard to know the dimension of the hidden ICs before. Secondly, it is dangerous to reduce the dimension using PCA in risk management as explained before. The criterion of PC choice depends on the explanation of the covariance matrix. However the risk measure such as Value-at-Risk (VaR) is the quantile in the tail that is measure by the 4-th moment. Therefore instead of reducing the dimension in the preprocess we choose m ICs according to their explanation of the tail afterwards. Another general model named "Noisy ICA" can be formulated as:

$$y(t) = Wx(t) + \varepsilon(t).$$

It is assumed that the noise $\varepsilon(t)$ is independent from ICs and normally distributed. This model not only compensates the assumption of the perfect linear relation but also relaxes the "nongaussianity" requirement discussed in the following section. However the covariance estimation of the normal error makes the model complicated. Normally the covariance is assumed to be known or equal to $\sigma^2\mathcal{I}$, Hyvärinen (1999). Compared to these generalizations, the basic ICA is computationally fast and reliable. Thus in our study we apply the basic ICA.

2.1 Definition and Properties

In this section, we will define the ICA model and discuss its properties. The basic ICA assumes that ICs are the linear transformation of the dependent observations. Due to the ambiguities of ICA discussed in a second, we denote the genuine independent sources as $s(t)$ and the ICs explained by the linear transformation as $y(t)$. Both notations refer to ICs but can have different orders and signs. Furthermore, it is found that ICA is infeasible for high-dimensional normal variables. The theoretical backgrounds of several ICA algorithms will be discussed then. Among them, FastICA is computationally efficient especially in a high-dimensional space, which will be intensively discussed in the next section.

2.1.1 ICA definition

Given a dependent vector $x(t) = \{x_1(t), \cdots, x_d(t)\}^\top$, we want to find a linear representation $x(t) = As(t)$ where $s(t) = \{s_1(t), \cdots, s_d(t)\}^\top$ is statistically independent. In other words, the joint density of $s(t)$ must equal the product of the marginal densities of its elements, see Definition 1.

$$\begin{aligned} \begin{pmatrix} x_1(t) \\ \vdots \\ x_d(t) \end{pmatrix} &= \begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \cdot & \cdots & \cdot \\ a_{d1} & \cdots & a_{dd} \end{pmatrix} \begin{pmatrix} s_1(t) \\ \vdots \\ s_d(t) \end{pmatrix} \\ \text{or equivalently, } x(t) &= As(t) \end{aligned} \tag{2.1}$$

where A is assumed to be a nonsingular matrix.
Alternatively, we can formulate ICA as:

$$\begin{aligned} \begin{pmatrix} y_1(t) \\ \vdots \\ y_d(t) \end{pmatrix} &= \begin{pmatrix} w_{11} & \cdots & w_{1d} \\ \cdot & \cdots & \cdot \\ w_{d1} & \cdots & w_{dd} \end{pmatrix} \begin{pmatrix} x_1(t) \\ \vdots \\ x_d(t) \end{pmatrix} \\ \text{or equivalently, } y(t) &= Wx(t) \end{aligned} \tag{2.2}$$

where $y(t) = \{y_1(t), \cdots, y_d(t)\}^\top$ is an independent vector and W is the inverse of A. We distinguish $s(t)$ and $y(t)$ due to the following ambiguities of ICA.

Firstly the variances of $s(t)$ and A (or $y(t)$ and W) are not defined uniquely since we estimate two terms $s(t)$ and A from one vector $x(t)$. For example, a scale increase of one term can be removed by the decrease of the other. Given a constant vector $c = (c_1, \cdots, c_d)^\top$, we get:

$$x_m(t) = \sum_j a_{mj} s_j(t) = \sum_j \{c_j a_{mj}\} \{\frac{1}{c_j} s_j(t)\}, \quad 1 \le m \le d, \quad j = 1, \cdots, d.$$

The term $\{cs(t)\}$ is obviously a new IC vector but its covariance is $(c\Sigma_s c^\top)$. Hence in ICA, $y(t)$ can be different from the real independent sources $s(t)$ by a scale. In order to avoid this ambiguity, it is convenient to specify the covariance of the independent sources to an identity matrix. The decomposition of the observed vector $x(t)$ is applied before ICA. This has no influence on the estimation result since the eigenvector and eigenvalue matrix are time invariant. This is called "prewhitening" process. After prewhitening, the matrix A becomes an orthogonal matrix, i.e. $A^\top A = \mathbf{I} = AA^\top$ where $\mathbf{I}$ denotes the identity matrix. It is same for its inverse matrix W. Remind that c_j can be negative, therefore the sign of IC is still not uniquely determined using prewhitening. Nevertheless financial time series are more or less symmetric, the sign of the risk has a trivial effect. Secondly, the order of

ICs is ambiguous. Given a permutation matrix P, we obtain:

$$x(t) = As(t) = AP^{-1}Ps(t).$$

It is clear that $Ps(t)$ consists of the original ICs but in a different order. In our study, the order ambiguity is obscure. Furthermore, ICA is infeasible for Gaussian random variables. Let $s(t)$ be a standardized and independent d-dimensional Gaussian distributed vector, the joint density is:

$$f(s) = |2\pi\mathbf{I}|^{-\frac{1}{2}} \exp\left(-\frac{s^\top s}{2}\right) = |2\pi\mathbf{I}|^{-1/2} \exp\left(-\frac{||s||^2}{2}\right)$$

where $||\cdot||$ is the norm of a vector. Given the matrix A in Equation (2.1) and based on the Jacobian transformation, we get:

$$f(x) = |2\pi\mathbf{I}|^{-\frac{1}{2}} \exp(-\frac{||A^\top x||^2}{2})|\det A^\top| = |2\pi\mathbf{I}|^{-1/2} \exp(-\frac{||x||^2}{2}).$$

The matrix A estimated in ICA disappears in the joint densities of $x(t)$. It implies that ICA doesn't change the distribution of the normal vectors. This effect can be explained by the property of the Gaussian variables. If they are uncorrelated, they are independent as well. Therefore, the prewhitening is enough to find the ICs of a dependent Gaussian vector. In our risk management study, the data are financial time series that follow heavy-tailed distributions, therefore the drawback of the basic ICA is negligible.

2.1.2 ICA algorithms and evaluation

Given the basic ICA model, there are 4 main ideas to find ICs. Here we report the evaluation of them to give an overview of the ICA development.

1. Maximization of nongaussianity is to find ICs which are maximally nongaussian. There is no assumption on the distribution of IC. Hyvärinen and Oja (1997) proposed a fast gradient algorithm based on this idea. It is applicable in a high-dimensional space.

2. Joint (approximate) decomposition of the 4-th order cumulant tensor is to find the matrix W by decomposing the cumulant groups using a common eigenvector. Cardoso and Souloumiac (1993) proposed an efficient algorithm in a space with a dimension lower than 40. But this algorithm has poor performance in a high-dimensional space.

3. Maximum likelihood estimation is to find ICs which maximizes the product of the marginal densities. It is attractive if the density of IC is given. Otherwise, an assumption is required in advance.

4. Minimization of mutual information has natural connection to the maximization of nongaussianity since both ideas are based on an information factor: entropy. Therefore, we eliminate the discussion here.

Maximization of nongaussianity

Based on the CLT, we can find ICs by maximizing nongaussianity of $\{Wx(t)\}$. The CLT says that a sum of independent variables is more normally distributed than any individual one. From Equation (2.1) and Equation (2.2) we can write:

$$y(t) = Wx(t) = WAs(t).$$

Let $Q = WA$ and q_{ij} denote the element of Q in the i-th row and j-th column with $i, j = 1, \cdots, d$. The above equation can be expressed as:

$$y_i(t) = \sum_{j=1}^{d} q_{ij} s_j(t).$$

According to the CLT, y_i is more Gaussian than any single IC s_j. The best estimate of $y_i(t)$ is obtained if one $q_{i,j}$ for $j = 1, \cdots, d$ equals ± 1 and the left 0. It implies that the best estimate compared to other alternatives is less gaussian. Hence one can estimate $y_i(t)$ which is maximally nongaussian. The objective function of the optimization depends on the measure of nongaussianity. The mainly used measures include kurtosis, 4-th order cumulant and negentropy.

DEFINITION 2 (Kurtosis and Excess Kurtosis):

$$\text{kurt}(y) = \frac{\mathsf{E}[(y-\mu_y)^4]}{\sigma^4} \tag{2.3}$$

$$\text{ekurt}(y) = \frac{\mathsf{E}[(y-\mu_y)^4]}{\sigma^4} - 3 \tag{2.4}$$

The objective function of ICA based on kurtosis is $\max\{\text{kurt}(y)\} = \{\text{kurt}(Wx)\}$. This optimization problem can be solved using a gradient method:

$$\frac{\partial|\text{kurt}(Wx)|}{\partial W}=0.$$

DEFINITION 3 (4-th Order Cumulant): Assume that y is a real-valued and continuous random vector with a zero mean, its first 4 order cumulants are:

$$\text{cum}_{y_i} = \mathsf{E}[y_i] \tag{2.5}$$

$$\begin{aligned}
\mathrm{cum}_{y_i,y_j} &= \mathsf{E}[y_i y_j] && (2.6)\\
\mathrm{cum}_{y_i,y_j,y_k} &= \mathsf{E}[y_i y_j y_k] && (2.7)\\
\mathrm{cum}_{y_i,y_j,y_k,y_l} &= \mathsf{E}[y_i y_j y_k y_l] - \mathsf{E}[y_i y_j]\,\mathsf{E}[y_k y_l] - \mathsf{E}[y_i y_k]\,\mathsf{E}[y_j y_l] - \mathsf{E}[y_i y_l]\,\mathsf{E}[y_j y_k] && (2.8)
\end{aligned}$$

We can use the empirical values to estimate the kurtosis and the 4-th order cumulant. These two measures are however sensitive to outliers. Thus a more robust measure, negentropy, is used as well.

DEFINITION 4 (Negentropy):

$$\begin{aligned}
J(y) &= H(y_{gauss}) - H(y) && (2.9)\\
H(y) &= -\int f_y(\eta) \log f_y(\eta) d\eta. && (2.10)
\end{aligned}$$

where $H(\cdot)$ is the entropy of a random variable. The term y_{gauss} is a Gaussian variable with the same covariance as variable y.

From the definition we see that negentropy is the difference of the entropies of two variables. Therefore, we discuss some properties of entropy here.

Given a continuous random vector $y = (y_1, \cdots, y_d)^\top$ with a joint probability density function (pdf) f_y, the differential entropy $H(y)$ is defined as:

$$H(y) = -\int f_y(\eta) \log f_y(\eta) d\eta. \qquad (2.11)$$

Entropy has some desirable properties. For example, given a constant value c and a matrix B, the following equations hold:

$$\begin{aligned}
H(y + c) &= H(y) && (2.12)\\
H(By) &= H(y) + \log|\det B| && (2.13)
\end{aligned}$$

If the vector $y \in R^d$ is independent, we have:

$$H(y) = \sum_{i=1}^{d} H(y_i) \qquad (2.14)$$

In the information theory, entropy is used to measure the "disorder" of one variable. The more disorder or say more random the variable is, the larger its entropy. It was proved that Gaussian variable has the largest entropy among all random variables of unit variance, Cover and Thomas (1991). This explains why the negentropy is a good measure of nongaussianity.

Negentropy is scale invariant. Given $h = My$ we have $\mathsf{E}[hh^\top] = M\Sigma M^\top$, further since:

$$H(y_{gauss}) = \frac{1}{2}\log|\det\Sigma| + \frac{d}{2}(1+\log 2\pi),$$

the negentropy of the new variable $J(h)$ is as same as $J(y)$:

$$\begin{aligned} J(My) &= H(My_{gauss}) - H(My) \\ &= \frac{1}{2}\log|\det\Sigma| + 2\frac{1}{2}\log|\det M| + \frac{d}{2}(1+\log 2\pi) - \{H(y) + log|\det M|\} \\ &= \frac{1}{2}\log|\det\Sigma| + \frac{d}{2}(1+\log 2\pi) - H(y) \\ &= H(y_{gauss}) - H(y) = J(y) \end{aligned}$$

Given the objective function in terms of negentropy $\max\{J(y)\} = \{J(Wx)\}$, one can use gradient methods to estimate W. Theoretically all these nongaussianity measures require the knowledge of the pdf, which makes the estimation computationally burdensome. In the recent, Hyvärinen (1998) proposed some reliable approximations of negentropy without estimating density function. Hyvärinen and Oja (1997) introduced a fast gradient algorithm named "FastICA", which extensively shortens the computational time. Moreover, the FastICA algorithm is very fast in a high-dimensional space. We will go to details in the next section.

Joint (approximate) decomposition of the cumulant tensor

As discussed above, the 4th-order cumulant $\mathrm{cum}(x_i, x_j, x_k, x_l)$ can be considered as a measure of nongaussianity. It is a four-dimensional $(d\times d\times d\times d)$ array and its usage in ICA was highly developed after the contribution of Cardoso and Souloumiac (1993). They proved that the linear transformation of the 4th-order cumulant includes $\{d\times(d+1)\}$ units of $d\times d$ matrices that inherit the whole information of the original cumulant matrix and can be decomposed jointly by a common eigenmatrix. The eigenmatrix is actually the estimate of A in Equation (2.1). This joint approximate decomposition of the eigenmatrices (JADE) algorithm is simple and even faster than the FastICA, if one considers a not too large dimensional data. Recall that this decomposition is analogous to the common principal component (CPC) analysis, see Flury (1998). As same as the CPC, the JADE suffers the inability of decomposition in a high-dimensional space. Therefore, it is unreliable if the data has a dimension larger than e.g. 40.

Let $x(t)$ be the prewhitened observed vector, from the ICA definition, we have

$$x(t) = As(t). \tag{2.15}$$

Since $x(t)$ is a prewhitened vector, A is an orthogonal matrix and $A = (a_1, \cdots, a_d)^\top$.

The linear transformation of the 4-th order cumulant $F(M)$ is defined as:

$$F_{i,j}(M) = \sum_{k,l=1}^{d} \text{cum}(x_i, x_j, x_k, x_l) m_{kl} \quad i,j,k,l = 1, \cdots, d, \tag{2.16}$$

where m_{kl} is a $d \times d$ matrix with 1 at the k-th row and l-th column and 0 otherwise. For each pair (i,j), $F_{i,j}(M)$ is a $d \times d$ matrix.

Remember that the cumulant has an additive property, i.e. the cumulant of the sum of independent variables equals a sum of their cumulants. Given an independent vector, we can only get a nonzero 4th order cumulant if the indices of the independent variables are unique. Straightforwardly, we obtain:

$$\begin{aligned}
F_{i,j} &= \sum_{k,l} \text{cum}(x_i, x_j, x_k, x_l) m_{kl} \\
&= \sum_{k,l} \text{cum}(\sum_o a_{io} s_o, \sum_p a_{jp} s_p, \sum_q a_{kq} s_q, \sum_r a_{lr} s_r) m_{kl} \\
&\overset{\text{o,p,q,r=v}}{=} \sum_{k,l,v} \text{kurt}(s_v) a_{iv} a_{jv} a_{kv} a_{lv} m_{kl} \\
&= \sum_v \text{kurt}(s_v) a_{iv} a_{jv} \{\sum_{k,l} a_{kv} a_{lv} m_{kl}\} \\
&= \sum_v \text{kurt}(s_v) a_{iv} a_{jv} a_v^\top a_v \\
&= \sum_v a_v^\top \{\text{kurt}(s_v) a_{iv} a_{jv}\} a_v
\end{aligned}$$

Equivalently we can formulate the above equation in a matrix notation:

$$F = A \Lambda_M A^\top \tag{2.17}$$

where Λ_M is a diagonal matrix. In practice, since the cumulant tensor is symmetric, $\frac{d \times (d+1)}{2}$ instead of $d \times d$ groups are decomposed jointly.

Maximum Likelihood Estimation

Maximum likelihood (ML) estimation is a fundamental method widely used in statistics. Let f_y denote the joint density of the ICs. From ICA model: $y(t) = Wx(t)$, it is easy to show that the joint density of x can be formulated as:

$$\begin{aligned}
f_x(x) &= |\det W| f_y(y) = |\det W| \prod_{i=1}^{d} f_i(y_i) \\
&= |\det W| \prod_{i=1}^{d} f_i(w_i x)
\end{aligned} \tag{2.18}$$

where $f_i(y_i)$ is the marginal density of y_i and $W = (w_1, \cdots, w_d)^\top$.

The likelihood function of T observations of x can be written as:

$$L(x|B) = \prod_{t=1}^{T} f_x\{x(t)\} = \prod_{t=1}^{T} [\prod_{i=1}^{d} f_i\{w_i x(t)\} |\det W|] \tag{2.19}$$

The log-likelihood is algebraically simple:

$$l(x|B) = \sum_{t=1}^{T} \sum_{i=1}^{d} \log f_i\{w_i x(t)\} + T \log |\det W| \tag{2.20}$$

If the marginal densities of ICs $f_i(y_i)$ are known in advance or the knowledge such as the distribution family of the marginal densities is given, the ML estimation is a good choice. Once again, we can use gradient methods to find the estimate of W and in turn the independent vector $y(t)$. Such knowledge is however not always available in financial study. For this reason, semiparametric and nonparametric density estimation methods can be used to specify the unknown densities. However they are computationally expensive if the dimension of the independent vector is high. Sometimes, we can assume a distribution family to the financial time series. However it is still computationally slow the distribution parameters need to be estimated simultaneously. Instead, we prefer a two-step procedure, i.e. estimating firstly the ICs without considering the distribution assumption and the parameters afterwards.

Minimization of mutual information

Mutual information is a useful tool to measure the independence in the information theory.

DEFINITION 5 (Mutual Information):
Mutual information I among d random variables y_i with $i = 1, \cdots, d$ is defined as:

$$I(y_1, \cdots, y_d) = \sum_{i=1}^{d} H(y_i) - H(y) \tag{2.21}$$

where $H(\cdot)$ denotes the entropy and $y = (y_1, \cdots, y_d)^\top$.

Mutual information is nonnegative and goes to zero if the random variables are independent. Its definition in terms of entropy gives a useful interpretation of a "distance". Recall that a distance between the joint distribution and the product of marginal distributions is a measure of independence. The "distance" measured in mutual information is analogous.

Remember that $y(t) = Wx(t)$, the mutual information of y can be formulated as:

$$I(y_1, \cdots, y_d) = \sum_{i=1}^{d} H(y_i) - H(Wx) = \sum_{i=1}^{d} H(w_i x) - H(x) - \log|\det W| \tag{2.22}$$

where $W = (w_1, \cdots, w_d)^\top$. Since y is expected to be independent, its mutual information must be close to 0. Therefore minimizing mutual information directs the way to find ICs. The objective function is $\min\{I(y)\} = \{\sum_{i=1}^{d} H(w_i x) - \log|\det W|\}$, which is exactly equal to the maximum of negentropy. Therefore, we eliminate the discussion of minimizing mutual information here.

2.2 Negentropy approximations and FastICA algorithm

Among the ICA algorithms, the combination of negentropy and FastICA is the most efficient method especially given a high-dimensional data. In this section, we will discuss the detail of this algorithm.

2.2.1 Negentropy approximations

From the definition of negentroy, we see that the pdfs of the ICs are required before the calculation.

$$\begin{aligned} J(y) &= H(y_{gauss}) - H(y) \\ H(y) &= -\int f_y(\eta) \log f_y(\eta) d\eta. \end{aligned}$$

Furthermore, the integral needs a large amount of computational time. Therefore, approximations of negentropy are preferable. The classic approximation of negentropy is to use the higher-oder cumulants. Comon (1994) gave an approximation of negentropy based on the polynomial expansions of Gram-Charlier and Edgeworth. It can be shown as:

$$J(y) \approx \frac{1}{12}\mathsf{E}[y^3]^2 + \frac{1}{48}\text{ekurt}(y)^2 \tag{2.23}$$

Recall that y is a prewhitened variable here. This approximation is simple but we need to estimate the higher-order moments such as kurtosis. Recall that kurtosis is very sensitive to outliers. A better approximation was proposed by Hyvärinen (1998). Compared to the polynomial approximation, the new one is simple and accurate.

Assume that the information of the random variable y with a density $f(y)$ is:

$$\int f(y) G_i(y) dx = c_i, \quad i = 1, \cdots, n \tag{2.24}$$

where $G_i(\cdot)$ is some nonquadratic function and n the number of the different G functions. From the above equation, we can find infinity estimates of $f(y)$ which fulfil Equation (2.24). In other words, the entropy of y is not well-defined from the given information. As discussed before, entropy reaches $-\infty$ in the limit and the largest value if the variable is Gaussian distributed with the same mean and variance as y. For this reason, the maximum entropy method is used so that we can get one determined estimate. Cover and Thomas (1991) have proved that the density with a maximum entropy has a form:

$$f(y) = H \exp\{\sum_{i=1}^{d} h_i G_i(y)\}, \tag{2.25}$$

where H and h_i are constants determined by c_i. Under some regular conditions such as y has a zero mean and unit variance and the assumption that $f(y)$ is not far from the standardized Gaussian due to the ground of the maximum entropy method, we can obtain:

$$\hat{f}(y) = \varphi(y)\{1 + \sum_{i=1}^{d} c_i G_i(y)\}, \tag{2.26}$$

where $\varphi(y)$ is the pdf of the standardized Gaussian r.v. and $c_i = \mathsf{E}[G_i]$ from Equation (2.24). Further, we can obtain the approximation of the differential entropy:

$$H(y) \approx -\int \hat{f}(y) \log \hat{f}(y) dy \approx H(y_{gauss}) - \frac{1}{2} \sum_{i=1}^{n} c_i^2. \tag{2.27}$$

Normally two G functions are used, assuming that G_1 is odd and G_2 even. The negentropy approximation can be formulated as:

$$J(y) \approx k_1[\mathsf{E}\{G_1(y)\}]^2 + k_2[\mathsf{E}\{G_2(y)\} - \mathsf{E}\{G_2(y_{gauss})\}]^2 \tag{2.28}$$

where k_1 and k_2 are positive constants in accordance with different functions G_i. For example, in one approximation, we choose:

$$G_1^a(y) = y \exp(-y^2/2) \tag{2.29}$$
$$G_2^a(y) = \exp(-y^2/2) \tag{2.30}$$

then $k_1 = 36/(8\sqrt{3} - 9)$ and $k_2^a = 1/(2 - 6/\pi)$. When we use another approximation where

$$G_1^b(y) = y \exp(-y^2/2) \tag{2.31}$$
$$G_2^b(y) = |y|, \tag{2.32}$$

we have $k_1 = 36/(8\sqrt{3} - 9)$ and $k_2^b = 24/(16\sqrt{3} - 27)$.

Figure 2.1 compares the true negentropy of a simulated mixture variable $y = p\mathrm{N}(0, 1) +$

$(1-p)\mathrm{N}(1,4)$ with $p \in [0,1]$ where $\mathrm{N}(m,v)$ denotes a Gaussian variable with a mean of m and a variance of v. It shows that given different value of p, the two approximations are very close to the true negentropy (straight line). In our study, we will use the approximation (2.28) to measure nongaussianity.

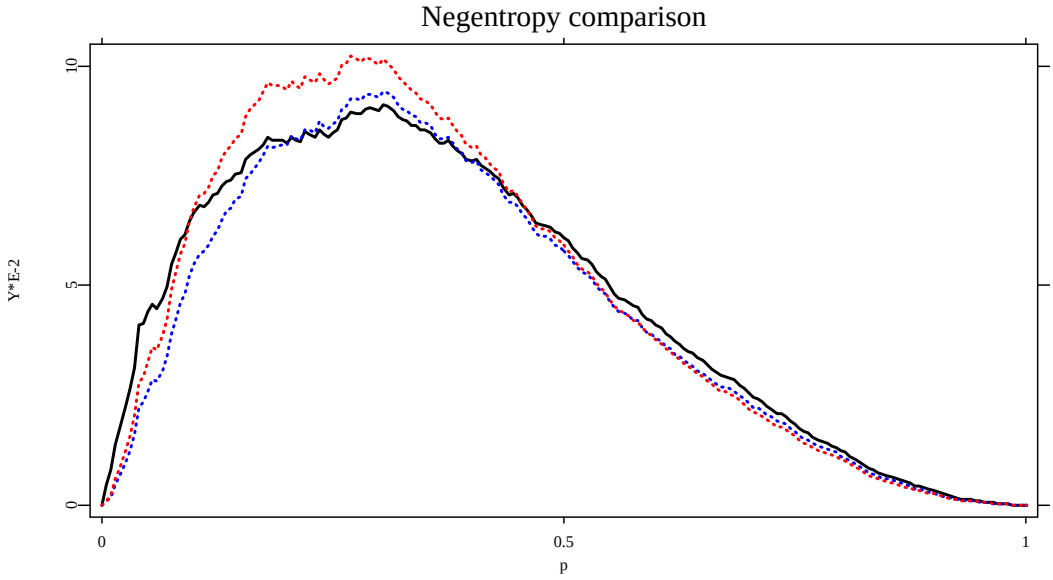

Fig. 2.1: Comparison of the true negentropy (solid) and its approximations (a: red and dashed, b: blue and dotted) of a simulated Gaussian mixture variable: $p\mathrm{N}(0,1) + (1-p)\mathrm{N}(1,4)$ for $p \in [0,1]$.

ICAnegentropyapp.xpl

2.2.2 *FastICA algorithm based on negentropy*

Given the negentropy approximations, the maximization of nongaussianity is an optimization problem. General gradient methods can find ICs with a low speed. Hyvärinen and Oja (1999) proposed a fast gradient method.

The objective function of ICA based on the negentropy approximation as the nongaussianity measure can be written as:

$$\begin{aligned} \max J(y|W) &\approx k_1[\mathsf{E}\{G_1(y)\}]^2 + k_2[\mathsf{E}\{G_2(y)\} - \mathsf{E}\{G_2(y_{gauss})\}]^2 \\ &= k_1[\mathsf{E}\{G_1(Wx)\}]^2 + k_2[\mathsf{E}\{G_2(Wx)\} - \mathsf{E}\{G_2(\mathrm{N}(0,1))\}]^2 \end{aligned} \tag{2.33}$$

Remember that y is prewhitened and hence the respective Gaussian variable is the standardized term $\mathrm{N}(0,1)$. The first term G_1 is an odd function and $k_1[\mathsf{E}\{G_1(y)\}]^2$ can vanish if y is symmetrically distributed. This is realistic in financial study where the time series are close to symmetry. This assumption induces an extreme simplification of the gradient

method. Therefore, an easier negentropy approximation can be formulated as:

$$\begin{aligned} J(y|W) &\approx k_2[\mathsf{E}\{G(y)\} - \mathsf{E}\{G(y_{gauss})\}]^2 \\ &\propto [\mathsf{E}\{G(Wx)\} - \mathsf{E}\{G(\mathrm{N}(0,1))\}]^2 \end{aligned} \tag{2.34}$$

where $G(\cdot)$ is an even function which was denoted as G_2 before. The application of the new negentropy approximation gave also good results even in the case that symmetry is not hold. Hyvärinen et al. (2001) proposed two choices of function G which have been proved useful:

Approximation 1:

$$G(y) = \frac{1}{a}\log\cosh ay, \quad 1 \leq a \leq 2 \tag{2.35}$$

$$g(y) \stackrel{\text{def}}{=} G'(y) = \tanh(ay) \tag{2.36}$$

$$g'(y) = a\{1 - \tanh^2(ay)\} \tag{2.37}$$

where $g(\cdot)$ denote the first derivative of G with respect to (w.r.t) y and $g'(\cdot)$ the second derivative of G. Very often, $a = 1$ is considered.

Approximation 2:

$$G(y) = -\exp(-y^2/2) \tag{2.38}$$

$$g(y) \stackrel{\text{def}}{=} G'(y) = y\exp(-y^2/2) \tag{2.39}$$

$$g'(y) = (1 - y^2)\exp(-y^2/2) \tag{2.40}$$

According to the classic gradient method, we get:

$$\{\mathsf{E}\{G(Wx)\} - \mathsf{E}[G\{\mathrm{N}(0,1)\}]\}\,\mathsf{E}\{xg(Wx)\}=0 \tag{2.41}$$

For the computational simplification, one gives an initial value of the first term $\omega = \{\mathsf{E}\{G(Wx)\} - \mathsf{E}[G\{\mathrm{N}(0,1)\}]\}$ due to the fact that it is close to some constant. The adaptive estimate of the first term can be obtained as:

$$\omega = G(Wx) - \mathsf{E}[G\{\mathrm{N}(0,1)\}]. \tag{2.42}$$

Note that the main information given by ω is the sign of the term. For example, if g is the tanh function, ω is -1 for supergaussian ICs.

Moreover the gradient method is speeded up including the constraint $||W||^2 = 1$. Then the gradient is formulated as:

$$\mathsf{E}\{xg(Wx)\} + \chi W = 0 \tag{2.43}$$

Solving the above equation by Newton's method, we get the iteration as:

$$W = W - \mathsf{E}[xg(Wx) + \chi W]/[\mathsf{E}\{g'(Wx)\} + \chi] \tag{2.44}$$

$$\text{hence:} \quad W = \mathsf{E}[xg(Wx) - \mathsf{E}\{g'(Wx)\}W] \tag{2.45}$$

In practice, there are two ways to estimate W. The first one accounts the vector w_i of W one by one and keeps the estimate uncorrelated from other $w_{j\neq d}$ with $i, j = 1, \cdots, d$. The second considers the d vectors simultaneously and orthogonalize the estimates together.

FastICA algorithm 1:

1. Set the number of ICs as d and $j = 1$.
2. Set $i = j$ and choose an initial vector w_i of unit norm.
3. Let $\tilde{w}_i = \mathsf{E}\{g(w_i x)x\} - \mathsf{E}\{g'(w_i x)\}w_i$, where g is the first derivative of G, g' the second derivative.
4. Othogonalization 1 (decorrelated): $\tilde{w}_i = w_i - \sum_{k\neq i}(w_i^\top w_k)w_k$
5. Othogonalization 2: $w_i = \tilde{w}_i/||\tilde{w}_i||$
6. If not converged, go back to 3.
7. Set $j = j + 1$. For $j \leq d$, go back to step 2.

FastICA algorithm 2 (symmetric orthogonalization):

1. Set the number of ICs as d.
2. Set $i = 1, \cdots, d$ and choose d initial vectors w_i of unit norm.
3. For each i, $\tilde{w}_i = \mathsf{E}\{xg(w_i^\top x)\} - \mathsf{E}\{g'(w_i^\top x)\}w_i$.
4. Symmetric othogonalization of $W = (w_1, \cdots, w_d)^\top$:

$$W = (WW^\top)^{-1/2}W$$

5. If not converged, go back to 3.

2.2.3 Dimension Reduction

As discussed before, we can choose the more important ICs and neglect the others. Here we suggest a measure to reduce the dimension of the original data. Briefly to say, we will consider m ICs and consequently the time-invariant matrix W is reduced to a $m \times d$ matrix.

Impelled by the idea in PCA, a measure of how well the first m PCs ordered by their variances explain variation is given by the relative proportion:

$$\varrho = \frac{\sum_{i=1}^{m} Var(PC_i)}{\sum_{i=1}^{d} Var(PC_i)}, \quad m \leq d, \tag{2.46}$$

we can analogously use a proportion ϱ to measure how well the first m ICs explain the tail of the joint distribution. Recall that the covariance of ICs is an identity matrix after the prewhitening. Since we use the negentropy to measure the tail, the relative proportion can be formulated as:

$$\varrho = \frac{\sum_{i=1}^{m} J(\tilde{y}_i)}{\sum_{i=1}^{d} J(\tilde{y}_i)}, \quad m \leq d, \tag{2.47}$$

where $\tilde{y}$ denote the ICs in a sequence ordered by their negentropies.

3. GENERALIZED HYPERBOLIC DISTRIBUTION

In Chapter 2 we introduced the fast gradient method to find independent components (ICs) $y(t)$ given the observed variables. The joint density of $y(t)$ is the product of the marginal densities of these ICs. The joint density of the original dependent variables $x(t)$ can be obtained based on the Jacobian transformation. In this sense, the problem to estimate their joint density is transformed to estimate their marginal densities in a univariate space. In this chapter, we are purpose to find good fits of the marginal densities of the ICs in financial markets. Here we simply assume that all the ICs follow one distribution family and estimate the unknown distribution parameters with respect to each one.

In the literature, for reasons of stochastic and numerical simplicity, it is often assumed that the involved risk factors are normally distributed. This is done e.g. in the RiskMetrics framework. However empirical studies have shown that financial risk factors have leptokurtic distributions which include a high peak and fat tails. Figure 3.1 displays the empirical distribution based on the daily standardized (devolatilized) returns of the foreign exchange (FX) rates DEM/USD from 1979/12/01 to 1994/04/01. It provides a strong evidence that the empirical density deviates from the normality assumption. The heavy-tailed feature is amplified when we compare the log densities of the kernel estimation and the normal fit. Although the nonparametric density estimation gives the most accurate fit in average, it has a poor performance in the tails where few observations are available. Another drawback of the nonparametric estimation comes from the fact that the quantile or the inverse of the density without form is more difficult to compute. On the contrary, parametric estimation gives the density form explicitly and is therefore preferable, if it captures the heavy tail property. There has been keen competition among various heavy-tailed distribution families. For example, conditional Gaussian distribution is able to mimic the fat tails well at a moderate VaR (e.g. 95%) confidence level. Nevertheless, the results are unsatisfactory for the extreme events such as profit and loss (P&L) at a 99% confidence level, Jaschke and Jiang (2002). In the recent, hyperbolic distributions have been attracting the attentions of researchers. Eberlein et al. (2003) applied the generalized hyperbolic (GH) distribution to the VaR calculation. Chen et al. (2005) calculated the VaR of the DEM/USD rates at extreme levels based on two density assumptions: Hyperbolic (HYP) and Normal-inverse Gaussian (NIG) distributions, two subclasses of the GH distribution. The backtestings of the two empirical studies showed that models with the GH distribution gave more accurate VaR values than that with the normal distribution.

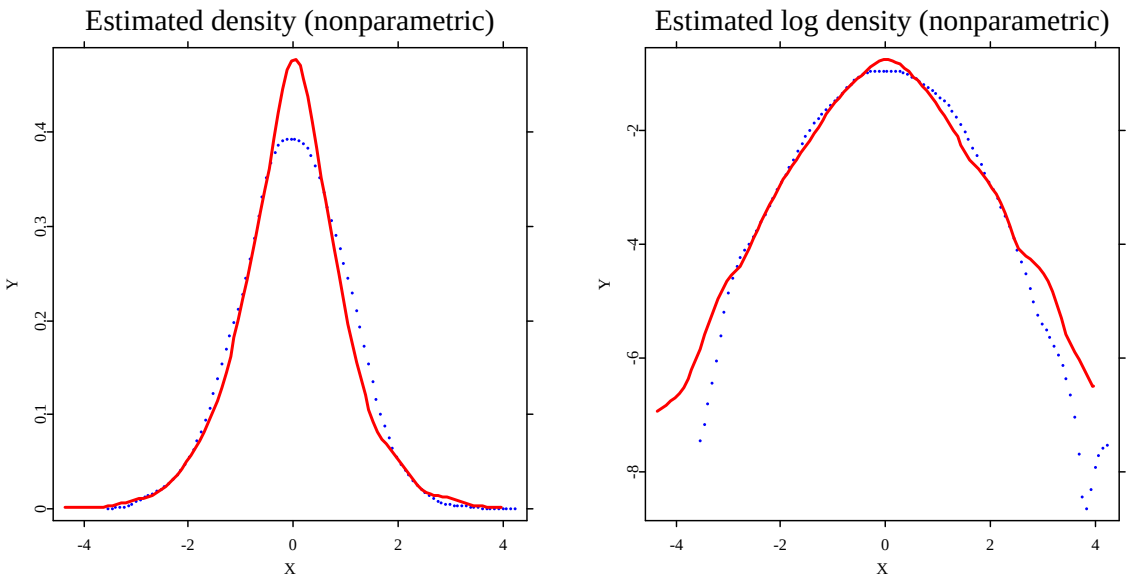

Fig. 3.1: Graphical comparison of the density (left) and the log-density (right) of the daily DEM/USD standardized returns from 1979/12/01 to 1994/04/01 (3719 observations). The kernel density estimate is graphed as a line and the normal density as dots with $h \approx 0.54$. Source: Chen et al. (2005).

It can be shown that the GH distribution is also a good assumption for the financial ICs. We consider the daily foreign exchange (FX) rates DEM/USD and GBP/USD from 1979-12-01 to 1994-04-01. Figure 3.2 compares the nonparametric density, the normal density and the hyperbolic (HYP) density estimations of the first and second ICs of the FX portfolio. It shows that the HYP fit almost coincides the nonparametric density. In the recent, copulas are used to fit the high-dimensional distribution, Embrechts, McNeil and Straumann (2002). Among them, t-copula has shown an outstanding performance. Therefore we also compared the t-distribution in the figure. Compared to the GH distribution, t-distribution has only one parameter: degree of freedom. However the over-fitting of the t distribution is often observed. For example, the t distribution with the degree of freedom 5 displays the very heavy tails in the marginal density of the first IC in Figure 3.2. For this reason, we prefer GH distribution to the t-distribution. It is however flexible to choose the suitable marginal distributions of the ICs based on the given data.

3.1 Univariate GH Distribution

The GH distribution introduced by Barndorff-Nielsen (1977) is a heavy-tailed distribution that can well replicate the empirical distribution of the financial risk factor. The density of

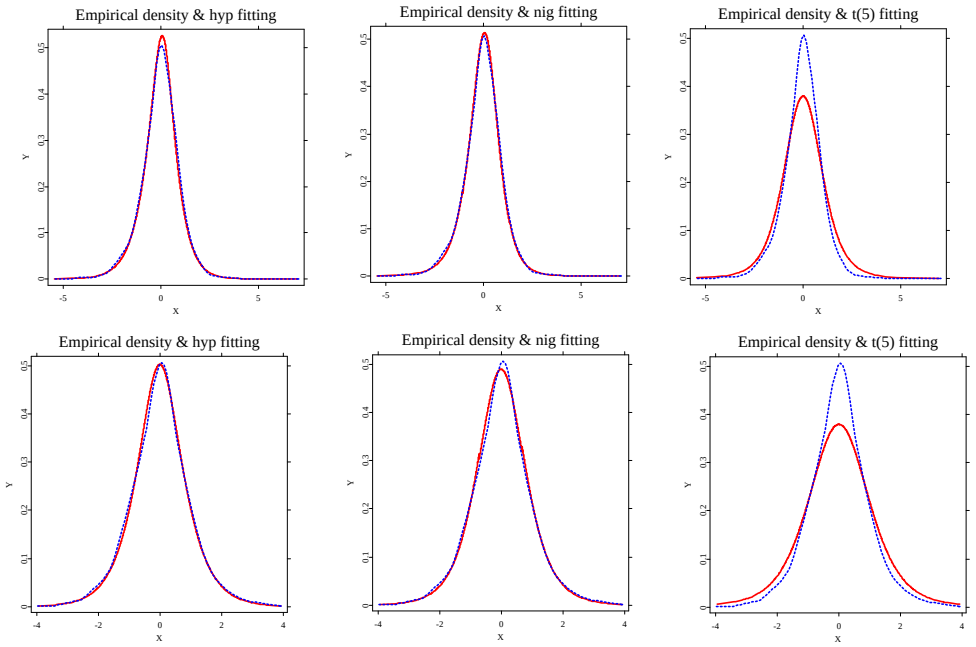

Fig. 3.2: Density estimations of the two ICs based on the daily DEM/USD and GBP/USD FX rates from 1979/12/01 to 1994/04/01 (3720 observations). The above from left are HYP, NIG and t(5) fittings respectively of the first IC. The bottom are the fittings of the second IC.
ICAfx.xpl

the GH variable $y \in \mathbb{R}$ is:

$$f_{GH}(y;\lambda,\alpha,\beta,\delta,\mu) = \frac{(\iota/\delta)^{\lambda}}{\sqrt{2\pi}K_{\lambda}(\delta\iota)} \frac{K_{\lambda-1/2}\left\{\alpha\sqrt{\delta^2+(y-\mu)^2}\right\}}{\left\{\sqrt{\delta^2+(y-\mu)^2}/\alpha\right\}^{1/2-\lambda}} \cdot e^{\beta(y-\mu)} \tag{3.1}$$

under the conditions:

- $\delta \geq 0$, $|\beta| < \alpha$ if $\lambda > 0$
- $\delta > 0$, $|\beta| < \alpha$ if $\lambda = 0$
- $\delta > 0$, $|\beta| \leq \alpha$ if $\lambda < 0$

where λ, α, β, δ and $\mu \in \mathbb{R}$ are the GH parameters, $\iota^2 = \alpha^2 - \beta^2$. The location and the scale of the density are mainly controlled by μ and δ respectively:

$$\begin{aligned} \mathsf{E}[X] &= \mu + \frac{\delta^2\beta}{\delta\iota}\frac{K_{\lambda+1}(\delta\iota)}{K_{\lambda}(\delta\iota)} \\ \mathrm{Var}[X] &= \delta^2\left\{\frac{K_{\lambda+1}(\delta\iota)}{\delta\iota K_{\lambda}(\delta\iota)} + (\frac{\beta}{\iota})^2[\frac{K_{\lambda+2}(\delta\iota)}{K_{\lambda}(\delta\iota)} - \{\frac{K_{\lambda+1}(\delta\iota)}{K_{\lambda}(\delta\iota)}\}^2]\right\}, \end{aligned}$$

whereas β and α play roles in the skewness and kurtosis of the distribution. For more details of the parameters' domains we refer to Bibby and Sørensen (2001). $K_\lambda(\cdot)$ is the modified Bessel function of the third kind with index λ, Barndorff-Nielsen and Blæsild (1981):

$$K_\lambda(y) = \frac{1}{2}\int_0^\infty y^{\lambda-1} exp\{-\frac{y}{2}(y+y^{-1})\}\, dy$$

Furthermore, the GH distribution has semi-heavy tails:

$$f_{GH}(y;\lambda,\alpha,\beta,\delta,\mu=0) \sim y^{\lambda-1}e^{-(\alpha-\beta)y} \text{ as } y\to\infty, \tag{3.2}$$

where $a(y) \sim b(y)$ as $y \to \infty$ means that both $a(y)/b(y)$ and $b(y)/a(y)$ are bounded as $y \to \infty$. Compared to the normal distribution, the GH distribution decays more slowly. However compared to the three heavy-tailed distributions: Student-t distribution, Laplace distribution and Cauchy distribution, the decaying speed of the GH distribution is often faster. The t-distribution can be formulated as:

$$f_t(y) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{\Pi\nu}}(1+\frac{y^2}{\nu})^{-(\frac{1+\nu}{2})},$$

for a large value of y, the density is proportional to $|y|^{-(\nu+1)}$, where ν is the degree of freedom.
The Laplace distribution is also called double exponential distribution with the form:

$$f_{Laplace} = \frac{1}{2\varsigma}e^{-|y-\mu|/\varsigma}$$

where μ is the location parameter and ς the scale parameter.
The Cauchy distribution is defined as:

$$f_{Cauchy} = \frac{1}{\varsigma\pi[1+(y-M)^2/\varsigma^2]}$$

where M is the median and ς the scale parameter.

In Figure 3.3 we compared these four heavy-tailed distributions with the normal one. In order to keep the comparability of these distributions, we specified the means of all the distribution to 0 and standardized their variances to 1. Instead of the GH distribution itself, we used one important subclasses of the GH distribution for simplicity: the normal-inverse Gaussian (NIG) distribution with $\lambda = -\frac{1}{2}$ introduced more precisely in the following text. On the left panel, the complete forms of these distributions are revealed. The Cauchy (dots) distribution has the lowest peak and the fattest tails, in other words, it has the flattest distribution. The NIG distribution decays second fast in the tails and has the highest peak. The tail behaviors are more clearly displayed on the right panel. Generally the GH distribution has an exponential decaying speed as shown in (3.2). By changing λ

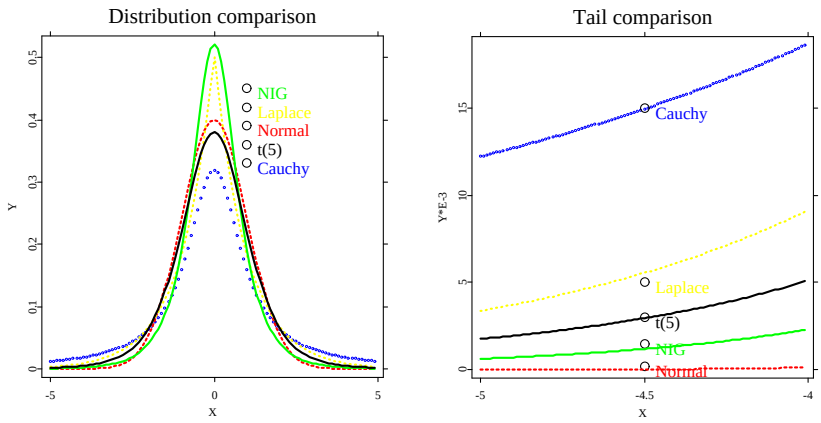

Fig. 3.3: Graphical comparison of the NIG distribution (line), standard normal distribution (dashed), Laplace distribution (dotted) and Cauchy distribution (dots).

ICAtail.xpl

the GH distribution family covers a wide range of tail behavior. For example, parameter $\lambda > 1$ introduces heavier tails than the double exponential. One may therefore think of λ as the tail control parameter, loosely speaking, with which one may model the range between a normal and a Cauchy tail.

The moment generating function of the GH distribution is:

$$m_f(z) = e^{\mu z} \cdot \frac{\iota^\lambda}{\iota_z^\lambda} \cdot \frac{K_\lambda(\delta \iota_z)}{K_\lambda(\delta \iota)}, \quad |\beta + z| < \alpha, \tag{3.3}$$

where $\iota_z^2 = \alpha^2 - (\beta + z)^2$. The GH distribution has the property that m_f is infinitely many times differentiable near 0, as a result every moment of a GH variable exists. In Section 2.2, this feature and the tail behavior (3.2) of the GH distribution will be used in the adaptive volatility estimation methodology.

In the current literature, subclasses of the GH distribution such as the hyperbolic (HYP) or the normal-inverse Gaussian (NIG) distribution are frequently used. This is motivated by the fact that the four parameters $(\mu, \delta, \beta, \alpha)^\top$ simultaneously control the four moment functions of the distribution, i.e. the trend, the riskiness, the asymmetry and the likeliness of the extreme events. Eberlein and Keller (1995), Barndorff-Nielsen (1997) have shown that these subclasses are rich enough to model financial time series and have the benefit of numerical tractability. Therefore in our study we concentrate ourselves on these two important subclasses of the GH distribution: HYP with $\lambda = 1$ and NIG distribution with $\lambda = -1/2$. The corresponding density functions are given as:

- Hyperbolic (HYP) distribution: $\lambda = 1$,

$$f_{HYP}(y; \alpha, \beta, \delta, \mu) = \frac{\iota}{2\alpha\delta K_1(\delta\iota)} e^{\{-\alpha\sqrt{\delta^2+(y-\mu)^2}+\beta(y-\mu)\}}, \tag{3.4}$$

  where $y, \mu \in \mathbb{R}$, $0 \leq \delta$ and $|\beta| < \alpha$,

- Normal-inverse Gaussian (NIG) distribution: $\lambda = -1/2$,

$$f_{NIG}(y; \alpha, \beta, \delta, \mu) = \frac{\alpha\delta}{\pi} \frac{K_1\left\{\alpha\sqrt{\delta^2+(y-\mu)^2}\right\}}{\sqrt{\delta^2+(y-\mu)^2}} e^{\{\delta\iota+\beta(y-\mu)\}}. \tag{3.5}$$

  where $y, \mu \in \mathbb{R}$, $\delta > 0$ and $|\beta| \leq \alpha$.

In order to estimate the unknown parameters $(\alpha, \beta, \delta, \mu)^\top$, the maximum likelihood (ML) and numerical optimization methods are used. For an i.i.d HYP resp. NIG distributed variable X, the log-likelihood functions are:

$$\begin{aligned} L_{HYP} &= T\log\iota - T\log 2 - T\log\alpha - T\log\delta - T\log K_1(\delta\iota) \\ &\quad + \sum_{t=1}^{T}\{-\alpha\sqrt{\delta^2+(y_t-\mu)^2} + \beta(y_t-\mu)\} \end{aligned} \tag{3.6}$$

$$\begin{aligned} L_{NIG} &= T\log\alpha + T\log\delta - T\log\pi + T\delta\iota \\ &\quad + \sum_{t=1}^{T}\left[\log K_1\left\{\alpha\sqrt{\delta^2+(y_t-\mu)^2}\right\} - \frac{1}{2}\log\{\delta^2+(y_t-\mu)^2\} + \beta(y_t-\mu)\right] \end{aligned} \tag{3.7}$$

Figure 3.2 shows the estimated HYP and NIG densities with the corresponding ML estimators of the two ICs of the FX portfolio. It can be seen that the estimated densities almost coincide with the empirical densities of the financial risk factors. The empirical density $\hat{f}_h(y)$ (line) was estimated by the kernel estimation:

$$\hat{f}_h(y) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{y-X_i}{h}), \tag{3.8}$$

where n is the number of observations and K is the kernel function which gives weights to the observations according to the distances of them to the fixed point y. The further an observation is from the fixed point, the smaller weight it will be given. We chose the Quartic kernel function with a closed form: $K(u) = \frac{15}{16}(1-u^2)^2\mathbf{1}(|u| \leq 1)$, where $\mathbf{1}(\cdot)$ is the indicator function which has a value of 1 if the condition in the parenthesis exists, and a value of 0 otherwise. In addition, we used the Silverman's rule of thumb to select the bandwidth h:

$$\hat{h}_{rot} \approx 1.06\hat{\sigma}n^{-1/5},$$

where $\hat{\sigma}$ is the empirical standard deviation of the variable X. Since the rule of thumb assumes that the unknown density belongs to the normal family and we chose the Quartic kernel, the bandwidth was adjusted to $\hat{h} \approx 2.62\hat{h}_{rot} = 2.78\hat{\sigma}n^{-1/5}$ using the canonical bandwidths. For details of the kernel and the bandwidth selections, see Chapter 3 in Härdle, Müller, Sperlich and Werwatz (2004). Compared to the normal distribution in Figure 3.2, it is convincing that the GH distribution family can represent the empirical distribution of financial risk very well.

3.2 Generating GH random variables

In this section, we discuss one algorithm to generate the GH random variable (r.v.). GH variable can be constructed by a generalized inverse Gaussian (GIG) random variable and a standard normal random variable, Cont and Tankov (2004). The GIG r.v. has a density function:

$$f_{GIG}(y) = \frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\psi\chi})} y^{\lambda-1} \exp\{-\frac{1}{2}(\chi y^{-1} + \psi y)\}, \quad y > 0,$$

where $\lambda \in R$ and χ, $\psi \in R^+$. The sampling of GH r.v. can be formulated as:

$$GH = \sqrt{GIG}\mathrm{N} + \beta GIG + \mu,$$

where N denotes the standardized normal r.v. and μ is a constant. Firstly, we discuss the algorithm to generate a GIG variable. Atkinson (1982) proposed an algorithm to generate the GIG r.v. using the envelope rejection technique. The density is assumed to be sampled by an exponential function. The range of the r.v. is divided into two parts and further a rejection function with respect to the value of the generated inverse of the GIG density is constructed. The algorithm to generate a $GH(\lambda, \alpha, \beta, \delta, \mu)$ r.v. is summarized as:

1. Set $\chi = \delta^2$ and $\psi = \alpha^2 - \beta^2$.
2. Generate a GIG r.v. $X \sim GIG(\lambda, \chi, \psi)$.
3. Generate a standard normal r.v. Z.
4. Get the GH r.v. $Y = \sqrt{X}Z + \beta X + \mu$.

4. ADAPTIVE VOLATILITY

In the heteroscedastic model $R_t = \Sigma_t^{1/2}\varepsilon_t$ the estimation of the covariance matrix Σ_t plays an important role. If the variables are independent, the covariance is a diagonal matrix. Thus we can estimate its elements in the diagonal line, i.e. the volatilities of ICs separately. The covariance of the original variables is a linear transformation of the diagonal matrix. Hence the covariance estimation is simplified to volatility estimation with respect to each IC. The development of volatility estimation can be summarized as:

$$\sigma \quad \rightarrow \quad \sigma_t = m(\cdot) \quad \rightarrow \quad \sigma_t = m_t(\cdot),$$

Briefly to say, volatility was firstly assumed to be constant, e.g. in the Black Scholes (BS) model introduced in 1973. This is however violated if the process spans a long time period. The so-called "volatility clustering" and the observed "volatility smile" in the option markets provided enough evidences to reject the constant assumption. Thereafter volatility is regarded to be time dependent. Firstly one used a time-invariant function m to estimate the volatility such as ARCH (Engle, 1995), GARCH (Bollerslev, 1995) and stochastic volatility models (Harvey et al., 1995). Such a model has an inherent weakness: For a long time series the form of the volatility model may be variant as the volatility itself with a very high possibility. It is therefore plausible to use a more flexible tool by providing a data-driven "local" model, which can avoid this potential miss-specification problem, i.e. $\sigma_t = m_t(\cdot)$ is used. Mercurio and Spokoiny (2003) introduced a model to estimate the volatility adaptively. The algorithm assumes that the volatility is constant in a short time interval although it can be volatile in a long run. The theory of the proposed "local constant" model was however based on the normality assumption of the risk factors. Chen et al. (2005) extended their study based on the GH distribution, which gave more accurate VaR estimations. In this chapter, we will introduce the adaptive estimation algorithm that will be applied to estimate the volatilities of the ICs.

4.1 Adaptive Volatility

The basic idea of adaptive volatility estimation comes from the observation that although the volatility is heteroscedastic in a long time period, its change in a short time interval, a so-called *time homogeneous interval*, is very small. Evidence for this argument has been

given by Mercurio and Spokoiny (2003). According to time homogeneity, one specifies an interval $I = [\tau - m, \tau)$ for a fixed time point τ with $0 \leq m \leq \tau - 1$, where the volatility σ_t, $t \in I$ is almost constant. One may for example estimate in this case the local constant volatility σ_τ by averaging the past squared returns r_t^2 for $t \in I$:

$$\hat{\sigma}_\tau^2 = \frac{1}{|I|} \sum_{t \in I} r_t^2, \tag{4.1}$$

where $|I|$ is the cardinality of I. Two questions arise in this procedure: how well does this estimate work and how to specify the homogeneous interval I?

The squared returns r_t^2 are always nonnegative and have for the stochastic errors ε_t (i.i.d. GH, HYP or NIG) a skewed distribution. In order to apply an interval selection procedure for I we use the power transformation of the return r_t:

$$|r_t|^\gamma = C_\gamma \sigma_t^\gamma + D_\gamma \sigma_t^\gamma \zeta_{\gamma,t} \tag{4.2}$$

where γ is the power transformation parameter, $\zeta_{\gamma,t} = (|\varepsilon_t|^\gamma - C_\gamma)/D_\gamma$ are standardized i.i.d. innovations, $C_\gamma = \mathsf{E}(|\varepsilon|^\gamma | \mathcal{F}_{t-1})$ is the conditional mean of $\zeta_{\gamma,t}$ and $D_\gamma^2 = \mathsf{E}[(|\varepsilon|^\gamma - C_\gamma)^2 | \mathcal{F}_{t-1}]$ is the conditional variance of $\zeta_{\gamma,t}$.

Additionally, lighter tails are obtained after this power transformation. Such a tail behavior is required later to derive theoretical properties of the estimate later. Let us denote the conditional mean of the transformed return $|r_t|^\gamma$ by θ_t:

$$\theta_t = C_\gamma \sigma_t^\gamma. \tag{4.3}$$

Since C_γ is a constant given a fixed γ, the estimate of the volatility σ_t is proportional to θ_t. In a time homogeneous interval I the local parameter θ_t is constant for $t \in I$ and is estimated by:

$$\hat{\theta}_I = \frac{1}{|I|} \sum_{t \in I} |r_t|^\gamma. \tag{4.4}$$

Writing $|r_t|^\gamma$ out in (4.4), one has:

$$\hat{\theta}_I = \frac{1}{|I|} \sum_{t \in I} \theta_t + \frac{s_\gamma}{|I|} \sum_{t \in I} \theta_t \zeta_t$$

where $s_\gamma = D_\gamma / C_\gamma$. One sees that the multiplicative error structure is turned via (4.2) into an additive one and the random variable $|r_t|^\gamma$ distributes more evenly. Straightforwardly, one can calculate the conditional expectation and variance of the estimate $\hat{\theta}_I$:

$$\mathsf{E}[\hat{\theta}_I | \mathcal{F}_{\tau-1}] \quad = \quad \mathsf{E} \frac{1}{|I|} \sum_{t \in I} \theta_t,$$

$$v_I^2 = \mathrm{Var}\,[\hat{\theta}_I|\mathcal{F}_{\tau-1}] \quad = \quad \frac{s_\gamma^2}{|I|^2}\,\mathsf{E}(\sum_{t\in I}\theta_t\zeta_t)^2 = \frac{s_\gamma^2}{|I|^2}\,\mathsf{E}\sum_{t\in I}\theta_t^2.$$

In a time homogeneous interval I, the volatilities are expected to be time invariant, therefore $\hat{\theta}_I$ can be considered as an estimate of θ_t for each time point $t \in I$. Therefore v_I can be estimated by:

$$\hat{v}_I = s_\gamma \hat{\theta}_I |I|^{-1/2}.$$

In other words, the volatility estimate σ_t can be induced from an estimate θ_t. However the specification of the local homogeneous interval is still open. Mercurio and Spokoiny (2003) have derived a homogeneity test for a supermartingale process. We show that a supermartingale of GH distributed variable can be obtained from the following lemma. It therefore leads to the same homogeneity test theory.

LEMMA 1: For every $0 \le \gamma \le 1$ there exists a constant $a_\gamma > 0$ such that

$$\log \mathsf{E}[e^{u\zeta_\gamma}] \le \frac{a_\gamma u^2}{2},$$

where $\zeta_\gamma = (|\varepsilon|^\gamma - C_\gamma)/D_\gamma$ is the transformed GH distributed variable ε.

The proof of this lemma is given in the Appendix.

Consider a predictable process p_t (such as the volatility σ_t or the local parameter θ_t) w.r.t. the information set $\mathcal{F}_{t-1}$:

$$\Upsilon_t = \exp\left(\sum_{s=1}^{t} p_s\zeta_s - (a_\gamma/2)\sum_{s=1}^{t} p_s^2\right)$$

Υ_t is a supermartingale, since

$$\begin{aligned}
\mathsf{E}(\Upsilon_t|\mathcal{F}_{t-1}) - \Upsilon_{t-1} &= \mathsf{E}(\Upsilon_t|\mathcal{F}_{t-1}) - \mathsf{E}(\Upsilon_{t-1}|\mathcal{F}_{t-1}) \\
&= \mathsf{E}[\exp\left(\sum_{s=1}^{t} p_s\zeta_s - (a_\gamma/2)\sum_{s=1}^{t} p_s^2\right) \\
&\quad - \exp\left(\sum_{s=1}^{t-1} p_s\zeta_s - (a_\gamma/2)\sum_{s=1}^{t-1} p_s^2\right)|\mathcal{F}_{t-1}] \\
&= \mathsf{E}[\exp\left(\sum_{s=1}^{t-1} p_s\zeta_s - (a_\gamma/2)\sum_{s=1}^{t-1} p_s^2\right)\{\exp(p_t\zeta_t - a_\gamma/2p_t^2) - 1\}|\mathcal{F}_{t-1}] \\
&= \underbrace{\frac{\exp(p_1\zeta_1)}{\exp(a_\gamma/2p_1)}}_{\le 1, Lemma 1}\cdots\underbrace{\frac{\exp(p_{t-1}\zeta_{t-1})}{\exp(a_\gamma/2p_{t-1})}}_{\le 1}\cdot\mathsf{E}[\underbrace{\frac{\exp(p_t\zeta_t)}{\exp(a_\gamma/2p_t)}}_{\le 1} - 1|\mathcal{F}_{t-1}] \\
&\le 0
\end{aligned}$$

i.e. $\mathsf{E}(\Upsilon_t|\mathcal{F}_{t-1}) \leq \Upsilon_{t-1}$. With this supermartingale property, the statistical properties of $\hat{\theta}_I$ are given in the following theorem.

THEOREM 1: If $R_1, ..., R_\tau$ obey the heteroscedastic model and the residual ε satisfies Lemma 1. Furthermore, the volatility coefficient σ_t satisfies the condition $b \leq \sigma_t^2 \leq bB$ with some positive constants b and B, then it holds for the estimate $\hat{\theta}_I$ of θ_τ :

$$P\{|\hat{\theta}_I - \theta_\tau| > \Delta_I(1 + \eta s_\gamma |I|^{-1/2}) + \eta \hat{v}_I\} \leq 4\sqrt{e}\eta(1 + \log B) \exp\{-\frac{\eta^2}{2a_\gamma(1 + \eta s_\gamma |I|^{-1/2})^2}\}.$$

where Δ_I is the squared bias defined as $\Delta_I^2 = |I|^{-1} \sum_{t \in I}(\theta_t - \theta_\tau)^2$.

Theorem 1 indicates that the estimation error $|\hat{\theta}_I - \theta_\tau|$ is small relative to $(\eta \hat{v}_I)$ for $\tau \in I$ with a high probability if I is a time homogeneous interval. Therefore the squared bias Δ_I is negligible. Straightforwardly, the following condition can be used to test the homogeneity hypothesis in an interval I:

$$|\hat{\theta}_I - \theta_\tau| \leq \eta \hat{v}_I.$$

In the test, I is split into two subintervals: $I \backslash J$ and J. If I is a time homogeneous interval, the estimates based on the two subintervals must be very close. The homogeneity condition can be stated as:

$$|\hat{\theta}_{I \backslash J} - \hat{\theta}_J| \leq \eta(\hat{v}_{I \backslash J} + \hat{v}_J) = \eta'(\sqrt{\hat{\theta}_J^2 |J|^{-1}} + \sqrt{\hat{\theta}_{I \backslash J}^2 |I \backslash J|^{-1}}). \tag{4.5}$$

provided $\eta' = \eta s_\gamma$ sufficiently large. If condition (4.5) is violated, the homogeneity hypothesis for the interval I is rejected.

The test procedure starts from an initial small interval I and consists of 4 steps:

Step 1: Enlarge the interval I from $[\tau - m_0, \tau)$ to $[\tau - k \times m_0, \tau)$, i.e. $m = k \times m_0$, and split the new interval into two subintervals J and $I \backslash J$. The parameters m_0 and k are integers specified according to data. In this paper, we chose $m_0 = 5$ and $k = 2$.

Step 2: Start homogeneity test for interval $J = [\tau - \frac{2}{3}m, \tau)$. If the homogeneity hypothesis isn't rejected, enlarge J one point further to $[\tau - \frac{2}{3}m - 1, \tau)$ and repeat the homogeneity test (4.5). The loop will continue until the left point of the subinterval J reaches the point $\tau - \frac{1}{3}m$. The choice of $\frac{1}{3}$ comes from the fact that the right $\frac{1}{3}$ part has been tested in the last homogeneous interval and the left one-thirds will be tested in the next homogeneous interval, Mercurio and Spokoiny (2003).

Step 3: If (4.5) is violated at point s, the loop stops and the time homogeneous interval I is specified from point τ to point $s + 1$.

Step 4: If time homogeneity holds for this interval, go back to Step 1.

The largest interval I is finally chosen as the time homogeneous interval for point τ, based on which the local volatility σ_τ is estimated. However there are still two threshold parameters to be specified: γ in the power transformation and η' in the homogeneity test condition. According to Lemma 1, the parameter γ is bounded in $[0, 1]$. In our study, we chose $\gamma = 0.5$ as same as the model based on the normal distribution to satisfy the comparability. The value of η' is similar to a smoothing parameter of the nonparametric regression. We thus propose a nonparametric way to pick up a global η'. Given a starting point t_0 and provided that there are enough past observations to estimate $\hat{\theta}_{(t_0,\eta')}$, the value η' minimizes the forecast error:

$$\eta' = \text{argmin} \sum_{t=t_0}^{\tau-1} \left\{ |r_t|^\gamma - \hat{\theta}_{(t,\eta')} \right\}^2. \tag{4.6}$$

4.2 Appendix

Proof of Lemma 1.

Proof:
Firstly we show that the moment generating function $\mathsf{E}[e^{u\zeta_\gamma}]$ exists for all $u \in R$.

Suppose that $\mathcal{L}(x) = GH(\lambda, \alpha, \beta, \delta, \mu)$ with the density function f for the transformed variable $y \stackrel{\text{def}}{=} |x|^\gamma$, we have

$$P(y \le z) = P(-z^{\frac{1}{\gamma}} \le x \le z^{\frac{1}{\gamma}}) = \int_{-\infty}^{z^{\frac{1}{\gamma}}} f(x)dx - \int_{-\infty}^{-z^{\frac{1}{\gamma}}} f(x)dx, \quad z > 0$$

Then the density of $y \in (0, \infty)$ is:

$$\begin{aligned} g(z) = \frac{d}{dz} P(y \le z) &= \gamma^{-1}\{f(z^{\frac{1}{\gamma}})z^{\frac{1}{\gamma}-1} + f(-z^{\frac{1}{\gamma}})z^{\frac{1}{\gamma}-1}\} \\ &= \gamma^{-1} z^{\frac{1}{\gamma}-1}\{f(z^{\frac{1}{\gamma}}) + f(-z^{\frac{1}{\gamma}})\}, \quad z > 0. \end{aligned}$$

Since $f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu = 0) \sim x^{\lambda-1}e^{-(\alpha-\beta)x}$ as $x \to \pm\infty$, it follows

$$\begin{aligned} g(z) &\sim \frac{z^{\frac{1}{\gamma}-1}}{\gamma}\{z^{\frac{\lambda-1}{\gamma}} e^{(\beta-\alpha)z^{\frac{1}{\gamma}}} + z^{\frac{\lambda-1}{\gamma}} e^{-(\beta+\alpha)z^{\frac{1}{\gamma}}}\} \\ &= \frac{z^{\frac{\lambda}{\gamma}-1}}{\gamma}\{e^{(\beta-\alpha)z^{\frac{1}{\gamma}}} + e^{-(\beta-\alpha)z^{\frac{1}{\gamma}}}\}, \quad z \longrightarrow \infty \end{aligned}$$

For $\gamma < 1$, it holds that $\int_0^\infty e^{uz} g(z)dz < \infty \ \forall u \in R$, since

$$\lim_{z\to\infty}(\beta - \alpha)z^{\frac{1}{\gamma}} + uz \quad \to \quad -\infty \quad \forall u \in R$$

$$\lim_{z\to\infty} -(\beta+\alpha)z^{\frac{1}{\gamma}} + uz \quad \to \quad -\infty \quad \forall u \in R$$

Since the integration depends only on the exponential part, it holds also that

$$\int_0^\infty z^n e^{uz} g(z)dz = \int_0^\infty \frac{\partial^n}{\partial u^n}(e^{uz})g(z)dz = \frac{\partial^n}{\partial u^n}\,\mathsf{E}[e^{uy}] < \infty,$$

then it can be shown that the moment generating function and $\log(\mathsf{E}[e^{uy}])$ are smooth. It holds for every $t > 0$,

$$\begin{aligned} \mathsf{E}[e^{uy}] = \mathsf{E}[e^{u|x|^\gamma}] &= \mathsf{E}[e^{u|x|^\gamma}\mathbf{1}(|x| \le t)] + \mathsf{E}[e^{u|x|^\gamma}\mathbf{1}(|x| > t)] \\ &\le e^{ut^\gamma} + \mathsf{E}[e^{|x|ut^{\gamma-1}} I(|x| > t)], \end{aligned} \tag{4.7}$$

Without loss of generality, we assume $\mu = 0$. Further

$$f_{GH}(x;\lambda,\alpha,\beta,\delta,\mu=0) \sim x^{\lambda-1}e^{-(\alpha-\beta)x} \text{ as } x \to \infty,$$

and $\int_y^\infty x^{\lambda-1}e^{-x}dx \sim y^{\lambda-1}e^{-y}$ as $y\to\infty$, Press, Teukolsky, Vetterling and Flannery (1992).

For an arbitrary but fixed $u \in R_+$ and $t_0 > 1$ so that $ut^{\gamma-1} < \alpha - \beta$, it holds for all $t \ge t_0$

$$\begin{aligned} f(t) &\le C_1 t^{\lambda-1}e^{(\beta-\alpha)t} \\ \int_{(\alpha-\beta-ut^{\gamma-1})t}^\infty x^{\lambda-1}e^{-x}dx &\le C_2[(\alpha-\beta-ut^{\gamma-1})t]^{\lambda-1}e^{-(\alpha-\beta-ut^{\gamma-1})t} \end{aligned}$$

where $C_1, C_2 > 1$.

Consequently for $t \ge t_0$,

$$\begin{aligned} \mathsf{E}[e^{u|t|^{\gamma-1}x}\mathbf{1}(|x| > t)] &= \int_t^\infty e^{ut^{\gamma-1}x} f(x)dx \le C_1 \int_t^\infty e^{ut^{\gamma-1}x} x^{\lambda-1}e^{-(\alpha-\beta)x}dx \\ &= C_1 \int_t^\infty x^{\lambda-1}e^{-(\alpha-\beta-ut^{\gamma-1})x}dx \\ &= C_1(\alpha-\beta-ut^{\gamma-1})^{-\lambda} \int_{(\alpha-\beta-ut^{\gamma-1})t}^\infty x^{\lambda-1}e^{-x}dx \\ &\le C_1C_2t^{\lambda-1}e^{-(\alpha-\beta-ut^{\gamma-1})t}(\alpha-\beta-ut^{\gamma-1}t)^{-1} \end{aligned} \tag{4.8}$$

If u is so large that $t \stackrel{\text{def}}{=} (\frac{\alpha-\beta}{2})^{\frac{1}{\gamma-1}}u^c \ge t_0$ with $\frac{1}{1-\gamma} \le c$, then (4.8) holds true since $ut^{\gamma-1} = (\frac{\alpha-\beta}{2})uu^{c(\gamma-1)} \le \frac{\alpha-\beta}{2} < \alpha-\beta$.

Given $t = (\frac{\alpha-\beta}{2}u)^{\frac{1}{1-\gamma}}$, we get

$$\mathsf{E}[e^{ut^{\gamma-1}x}\mathbf{1}(|x| > t)] \le \frac{2C_1C_2}{\alpha-\beta}(\frac{\alpha-\beta}{2}u)^{\frac{\lambda-1}{1-\gamma}}e^{-\frac{\alpha-\beta}{2}(\frac{\alpha-\beta}{2}u)^{\frac{1}{1-\gamma}}}.$$

From which we get

$$\log(\mathsf{E}[e^{ut^{\gamma-1}}\mathbf{1}(x>t)]) \leq C_3 + \frac{\lambda-1}{1-\gamma}\log(u) - (\frac{\alpha-\beta}{2})^{\frac{2-\gamma}{1-\gamma}}u^{\frac{1}{1-\gamma}}$$

Further $\log(\mathsf{E}[e^{ut^{\gamma-1}}\mathbf{1}(x > t)])u^{-\frac{1}{1-\gamma}}$ is also bounded for $u \to \infty$. Analogously we can show the bounding of $\log(\mathsf{E}[e^{ut^{\gamma-1}}\mathbf{1}(x < -t)])u^{-\frac{1}{1-\gamma}}$. Therefore for $\gamma < 1$ the whole term $\mathsf{E}[e^{u|x|^\gamma}\mathbf{1}(|x| > t)]u^{-\frac{1}{1-\gamma}}$ is bounded as $u \to \infty$.

Given $t = (\frac{\alpha-\beta}{2}u)^{\frac{1}{1-\gamma}}$, we have

$$\begin{aligned} e^{ut^\gamma} &= e^{(\frac{\alpha-\beta}{2})^{\frac{\gamma}{1-\gamma}}u^{\frac{1}{1-\gamma}}} \\ u^{-\frac{1}{1-\gamma}}\log(e^{ut^\gamma}) &= (\frac{\alpha-\beta}{2})^{\frac{\gamma}{1-\gamma}} = \text{constant} \end{aligned}$$

Thus $u^{-\frac{1}{1-\gamma}}\log(\mathsf{E}[e^{u|x|^\gamma}]) \leq u^{-\frac{1}{1-\gamma}}[\log(e^{ut^\gamma}) + \log\{E[e^{ut^{\gamma-1}|x|}\mathbf{1}(|x| > t)]\}]$ is bounded for $u \to \infty$, i.e. for a sufficient large u_0 there exist a constant $C_u > 0$ such that

$$\mathsf{E}[e^{u|x|^\gamma}] \leq C_u u^{\frac{1}{1-\gamma}}, \quad u \geq u_0.$$

□

5. EMPIRICAL STUDY

In this chapter, we will use the proposed methods: independent component analysis (Chapter 2), generalized hyperbolic (GH) distribution (Chapter 3) and adaptive volatility estimation (Chapter 4) to calculate the multivariate value at risk (VaR) of one portfolio: 2-dimensional exchange rates DEM/USD and GBP/USD.

VaR is one of the most often used risk measure. Given the distribution of the portfolio return, VaR is the quantile q_a. It measures the possible loss level over a given horizon at a given confidence level $1-a$ and answers the question: How much can I lose with a probability over the pre-set horizon. The research on VaR models has been ignited and prompted by the rule of Basel Committee on Banking Supervision in 1995: financial institutions may use their internal VaR models. The selection of the internal VaR model as well as the volatility estimation is essential to the VaR based risk management. Mathematically the VaR is defined as:

$$\mathrm{VaR}_{a,t} \stackrel{\mathrm{def}}{=} F_t^{-1}(a).$$

where F_t^{-1} is the inverse function of the conditional cumulative distribution function of the returns of the portfolio at time t, Franke et al. (2004). In the moment, three methods are mainly used in practice to measure the multivariate VaR:

- RiskMetrics, also named "Delta-Gamma-Normal" method. It assumes that the elements of the portfolio follow the Gaussian distribution and so does the portfolio itself. Given the weight of the elements $b = (b_1, \cdots, b_d)^\top$, we can forecast the VaR at the level a as:

$$\mathrm{VaR}_{a,t+1} = q_a \sqrt{x_t^\top \Sigma_t x_t}, \tag{5.1}$$

  where Σ_t is estimated using the exponential moving average (EMA) method, see Franke et al. (2004). RiskMetrics is definitely easy to implement with the desirable statistical properties of the normal distribution. Asymptotically, if the dimension of the portfolio is large enough, the portfolio converges to the Gaussian distribution. It concurs with the diversification effect of a large portfolio. In this sense, RiskMetrics method gives an approximation of the quantile of the portfolio return. However this method often underestimates the market risk of the portfolio since the most financial time series are heavy-tailed distributed. Since RiskMetrics is widely used as a standard in the financial industry, we will compare the accuracy of the VaR forecasts based on

the proposed methododlogy and the RiskMetrics one.

- Monte Carlo simulation gives the most precise VaR calculation. It simulates the time series of each element included in the portfolio and estimate the covariance matrix based on the simulations. However, the implementation of this method is jeopardized by its expensive computation.

- Historical simulation claculates the VaR by constructing a univariate hypothetical portfolio return. It therefore avoids the estimation of covariance and is possibly immune against a fault-pron distribution assumption. Given the univariate time series, we can find a good fitting for its empirical distribution. For example, a nonparametric fitting or a heavy-tailed distribution fitting. It however losses the detailed information on how risky is each element. Moreover a new hypothetical series needs to be constructed again if the weights of the portfolio are changed. This is not the case for the RiskMetrics, Monte Carlo simulation method and neither for our method proposed.

Given a d-dimensional portfolio, we propose a VaR calculation process as:

1. Use FastICA to get d ICs. If d is too large, the dimension reduction based on the measure ϱ can be considered, see Equation (2.47). Suppose now m ICs are chosen.

2. Estimate the adaptive volatility process for each IC. Since the volatility process is a supermartingale, it is natural to use the estimate of today as the forecast of tomorrow $\hat{\sigma}_{t+1}$.

3. Assume that each IC follows the GH distribution (subclasses: HYP or NIG) and calibrate the distribution parameters respectively. Here instead of calibrating IC series, we handle the devolatilized IC: $y(t)/\sigma_t$ which is comparably auto-uncorrelated.

4. Apply Monte Carlo (MC) simulation to estimate the quantile q_a of the portfolio. Given the weighting vector $b = (b_1, \cdots, b_d)^\top$, the portfolio return can be predicted as:

$$\begin{aligned}
\hat{R}_{t+1} &= b^\top \hat{x}_{t+1} \\
&= b^\top \begin{pmatrix} w_{11} & \cdots & w_{m1} \\ \vdots & \vdots & \vdots \\ w_{1d} & \cdots & w_{md} \end{pmatrix} \begin{pmatrix} \hat{y}_1(t+1) \\ \vdots \\ \hat{y}_m(t+1) \end{pmatrix} \\
&= b^\top \begin{pmatrix} w_{11} & \cdots & w_{m1} \\ \vdots & \vdots & \vdots \\ w_{1d} & \cdots & w_{md} \end{pmatrix} \begin{pmatrix} \hat{\sigma}_{1,t+1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{\sigma}_{m,t+1} \end{pmatrix} \begin{pmatrix} \hat{\varepsilon}_{1,t+1} \\ \vdots \\ \hat{\varepsilon}_{m,t+1} \end{pmatrix} \\
&= b^\top W^\top \hat{\Sigma}_{t+1}^{1/2} \hat{\varepsilon}_{t+1}
\end{aligned} \tag{5.2}$$

For each independent stochastic variable ε_i with $1 \leq i \leq m$, we simulate N times GH distributed variables with T observations given the estimated distribution parameters. The average quantile $\bar{q}_a = \sum_{i=1}^{N} q_a$ of the portfolio return is considered as the VaR forecast. Note that the weight b can be time variant. However this adjustment will not influence the estimations of ICs and further their marginal densities and volatilities.

To evaluate the validation of the VaR calculation, we consider the backtesting procedures presented in Christoffersen (1998). Since RiskMetrics is regarded as an industrial standard measure of portfolio risk, we will compare the forecasts of the proposed method with those based on the RiskMetrics method. Above all, a reliable VaR calculation should not underestimate the market risk. Let $\mathbf{1}_t$ denote the indicator of exceptions at time point t, $t = 1, 2, \ldots, \tau$, when the true loss is over the VaR forecast. If the proportion of exceptions, $n/\tau = \tau^{-1} \sum_{t=1}^{\tau} \mathbf{1}_t$, is much larger than a, it means that the possible losses happen more often than the fixed level. In this case, the VaR model should be rejected. One constructs hypotheses as:

$$\begin{aligned} H_0 &: \mathsf{E}[n] = \tau a \\ H_1 &: \mathsf{E}[n] \neq \tau a \end{aligned}$$

Under H_0, n is a Binomial random variable with parameters τ and a, the likelihood ratio test statistic can be derived as:

$$\mathrm{LR1} = -2 \log \{(1-a)^{\tau-n} a^n\} + 2 \log \{(1-n/\tau)^{\tau-n} (n/\tau)^n\}, \tag{5.3}$$

which is asymptotically $\chi^2(1)$ distributed, Jorion (2001).

In addition, a VaR model that yields exception clusters should also be rejected. A cluster of VaR exceedances means that if there is an exception today, an exception may also occur tomorrow with a higher probability than the prescribed level a. Therefore the second backtesting is the test of independence. Let us denote $\pi_{ij} = P(\mathbf{1}_t = j | \mathbf{1}_{t-1} = i)$ as the transition probability and $n_{ij} = \sum_{t=1}^{\tau} \mathbf{1}(\mathbf{1}_t = j \text{ and } \mathbf{1}_{t-1} = i)$, where $i, j = 0$ or 1. The independence null hypothesis is given as:

$$H_0 : \pi_{00} = \pi_{10} = \pi, \pi_{01} = \pi_{11} = 1 - \pi \tag{5.4}$$

One can test this hypothesis using the likelihood ratio statistic:

$$\mathrm{LR2} = -2 \log \{\hat{\pi}^{n_0} (1-\hat{\pi})^{n_1}\} + 2 \log \{\hat{\pi}_{00}^{n_{00}} \hat{\pi}_{01}^{n_{01}} \hat{\pi}_{10}^{n_{10}} \hat{\pi}_{11}^{n_{11}}\}, \tag{5.5}$$

where $\hat{\pi}_{ij} = n_{ij}/(n_{ij} + n_{i,1-j})$, $n_j = n_{0j} + n_{1j}$, and $\hat{\pi} = n_0/(n_0 + n_1)$. Under H_0, LR2 is asymptotically $\chi^2(1)$ distributed as well, Jorion (2001).

Time Series	Mean	Variance	Skewness	Kurtosis	ρ_1	ρ_2
DEM/USD	-4.71e-06	5.03e-05	-0.13	4.94	0.02	0.01
GBP/USD	0.00	4.86e-05	-0.01	5.64	0.08	0.01
IC1	-0.02	1.00	-0.62	8.71	0.07	0.02
IC2	0.01	1.00	-0.08	5.19	0.05	0.01

Tab. 5.1: Descriptive statistics of the returns of DEM/USD and GBP/USD and their two ICs.
ICAfxdescriptive.xpl

5.1 Exchange Rates: DEM/USD and DEM/JPD

The first portfolio we considered consists of two daily exchange rates: DEM/USD and GBP/USD from 1979-12-01 to 1994-04-01. Each series has 3721 observations. The data is available at MD*Base (www.mdtech.de). We firstly use FastICA to find two ICs based on the daily returns of the exchange rates. From the descriptive statistics reported in Table 5.1, we can see that all the four time series: returns of DEM/USD and GBP/USD and the two ICs are standardized, symmetric and leptokurtic. The distribution assumptions made in our study are then sound. Moreover the autocorrelations (ρ_1 and ρ_2) are insignificant. The dependent structure of the two time series is illustrated in Figure 5.1. It displays a strong positive linear relation between the two return processes in the left plot. This dependence is then eliminated by the linear transformation of ICA. On the right panel of this figure, the two ICs seems linearly uncorrelated. A more convincing result is illustrated in the comparison of the joint densities of the returns of the two exchange rates. Figure 5.2 shows the nonparametrical kernel density estimation on the left and the product of the HYP fittings of two ICs on the right. Firstly, the nonparametrically estimated joint density of the returns (red surface) deviates from the RiskMetrics assumption, the bivariate normal density (blue surface). As emphasized before, the tail of the joint density is crucial in the VaR calculation. The nonparametrically density has a higher peak and more important heavier tails than the normal one. Therefore in the RiskMetrics framework, one often underestimates the probability of the extreme values. On the right panel, the product of the marginal densities of the ICs imitates the whole surface of the empirical joint density, especially in the tails. In this case, it provides an evidence that the proposed methodology is better.

The time series of the two ICs are obtained based on the ICA model. The estimates of the matrices in ICA are:

$$\hat{W} = \begin{pmatrix} 207.93 & -213.63 \\ 77.72 & 73.29 \end{pmatrix} \quad \hat{A} = \begin{pmatrix} 2.30e-3 & 6.71e-3 \\ -2.44e-3 & 6.53 \end{pmatrix} \tag{5.6}$$

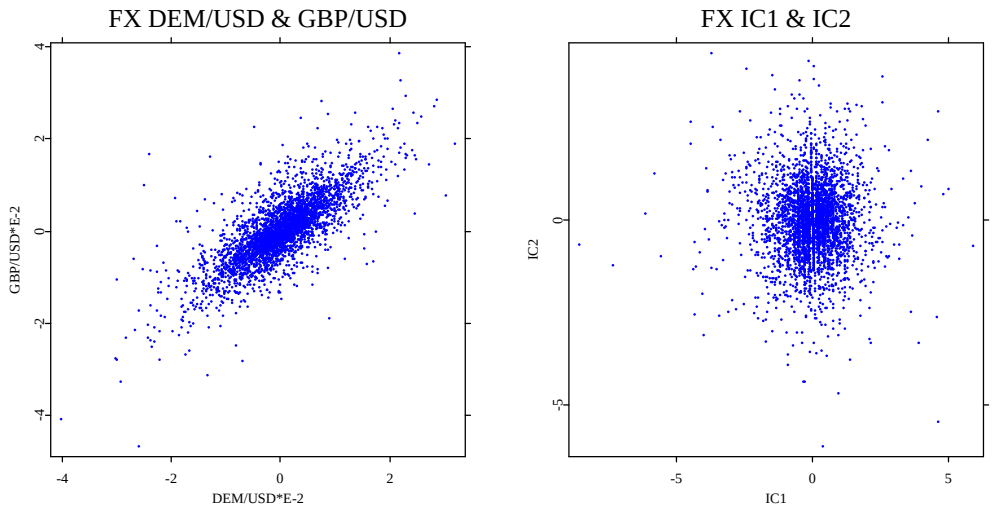

Fig. 5.1: FX returns DEM/USD and GBP/USD and the ICs of the FX returns.
ICAfxdescriptive.xpl

For each IC, we use the adaptive constant model to estimate the time-variant volatility process. Figure 5.4 displays the time series of the two adaptive volatility processes. It is interesting to observe the seasonality in the volatility process of the second IC. It implies that one risk factor behaves more volatile around summer regularly. As discussed before, the devolatilized return is expected to be more uncorrelated. The ACF plots of the devolatilized IC processes are shown in Figure 5.3. The two processes are both auto-uncorrelated. Based on the devolatilized return processes we estimate the GH distribution parameters.

In our study, we consider four trading strategies with $b_1 = (1,1)^\top$, $b_2 = (1,2)^\top$, $b_3 = (-1,2)^\top$ and $b_4 = (-2,1)^\top$. The VaR forecasts of five extreme risk levels of the portfolios at $a = 5\%$, 1%, 0.5%, 0.25% and 0.1% are forecasted and backtested for the last 4 years (1000 forecasts). The VaR time plots with exceptions are displayed in Figure 5.5 to Figure 5.8. The portfolio returns are plotted in dots and the 1000 one-step-ahead VaR forecasts based on the proposed methodology are drawn in a line. The forecasts based on the RiskMetrics are plotted in a dotted line. Exceptions when the portfolio losses more than the VaR forecast are marked as a cross. The crosses above denote the exceptions in the proposed methodology and those on the bottom are the exceptions observed in the RiskMerics framework. The empirical risk levels and the corresponding backtesting results w.r.t. different trading strategies are reported in Table 5.2. For different trading strategies and fixed VaR risk level, we calculated the empirical risk level, i.e. the proportion of the exceptions, the values of the two tests: LR1 and LR2. The RiskMetrics underestimates the market risks and the level tests are rejected. The worst case happens when one holds 1 unit DEM/USD and 2 units GBP/USD in his portfolio, the empirical risk level is 25 times

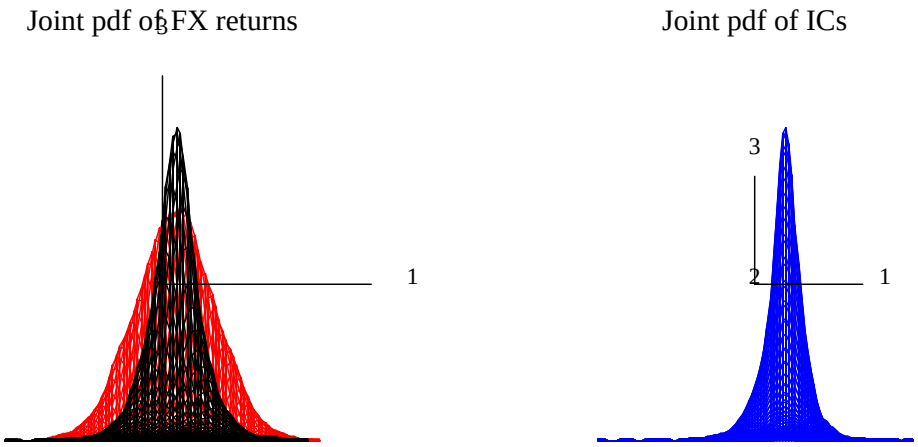

Fig. 5.2: Comparison of the nonparametrical joint density (black) of the returns of the exchange rates and the product (blue) of the HYP marginal densities of two ICs. The red surface is the Gaussian fitting with the same covariance as the returns of the exchange rates.

ICAfxjointdensitycomp.xpl

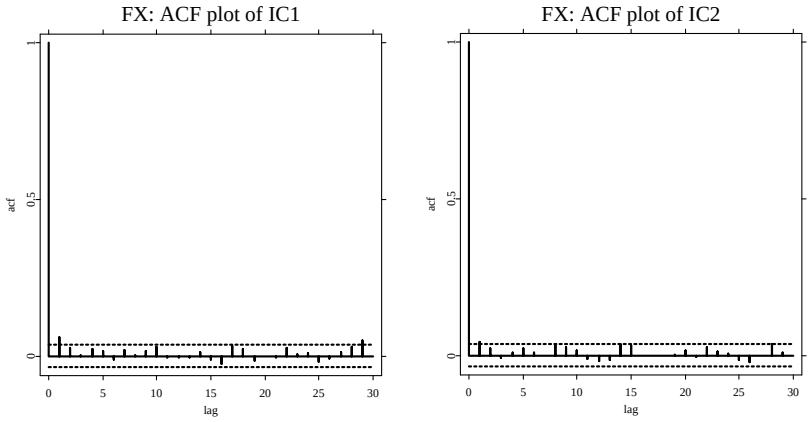

Fig. 5.3: ACF plots of the devolatilized ICs on the base of DEM/USD and GBP/USD rates.

ICAfxdescriptive.xpl

of the expected one (0.1%). On the contrary, the proposed method provides very accurate VaR forecasts. The empirical level is not only underestimated but very close to the level expected. This benefits the financial institutions who want to control their risks and at the same time avoid unnecessary capital deposit. However, we observe that the proposed VaR model gives too deep values, especially after a large loss. The more extreme the risk level is, the more turbulent the VaR forecast.

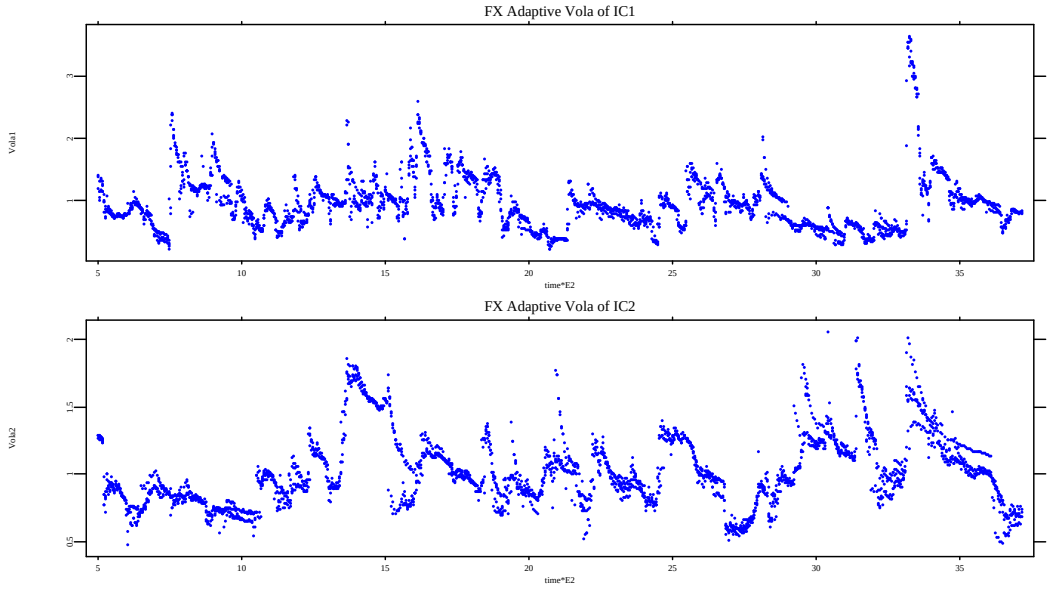

Fig. 5.4: Adaptive volatility time series of the ICs of the FX returns.

ICAfxdescriptive.xpl

		Proposed Methodology			RiskMetrics Method		
$b^\top$	$a(10^-2)$	$n/\tau(10^-2)$	LR1	LR2	$n/\tau(10^-2)$	LR1	LR2
(1,1)	5.00	4.5	0.54	31.91	10.8	53.96	19.85
(1,1)	1.00	0.9	0.10	0.00	5.5	99.60	21.71
(1,1)	0.50	0.2	2.34	0.00	5.0	142.33	23.44
(1,1)	0.25	0.2	0.11	0.00	3.8	137.10	21.05
(1,1)	0.10	0.1	0.00	0.00	2.3	100.72	27.74
(1,2)	5.00	4.1	1.81	27.22	10.9	55.64	24.79
(1,2)	1.00	0.9	0.10	0.00	5.3	92.67	22.38
(1,2)	0.50	0.4	0.22	0.00	4.7	128.43	21.78
(1,2)	0.25	0.2	0.11	0.00	4.0	148.23	24.41
(1,2)	0.10	0.0	2.00	-NAN	2.5	113.53	18.06
(-1,2)	5.00	3.9	2.75	29.63	12.0	75.40	35.67
(-1,2)	1.00	1.0	0.00	5.35	6.0	117.58	25.86
(-1,2)	0.50	0.6	0.19	9.11	4.4	114.93	18.33
(-1,2)	0.25	0.2	0.11	0.00	3.1	99.92	21.27
(-1,2)	0.10	0.0	2.00	-NAN	2.5	113.53	23.82
(-2,1)	5.00	3.6	4.55	11.89	12.7	89.19	18.41
(-2,1)	1.00	0.8	0.43	0.00	5.6	103.12	14.79
(-2,1)	0.50	0.4	0.22	0.00	4.2	106.16	2.41
(-2,1)	0.25	0.1	1.17	0.00	3.2	105.05	5.54
(-2,1)	0.10	0.0	2.00	-NAN	2.3	100.72	6.22

Tab. 5.2: Backtesting of the VaR forecast of the exchange portfolios.

ICAfxVaRplot.xpl

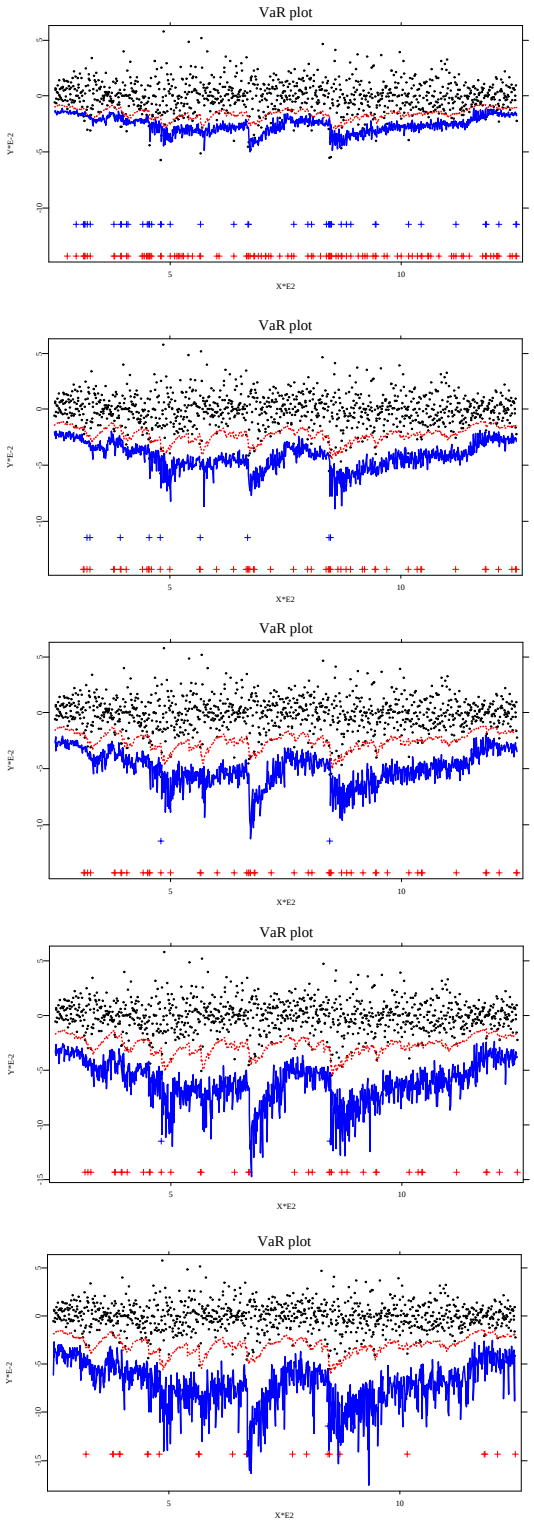

Fig. 5.5: VaR time plots of the exchange rate portfolio with the weight $b = (1, 1)^\top$. The risk levels are 95%, 99%, 99.5%, 99.75% and 99.9% respectively from the top.

ICAfxVaRplot.xpl

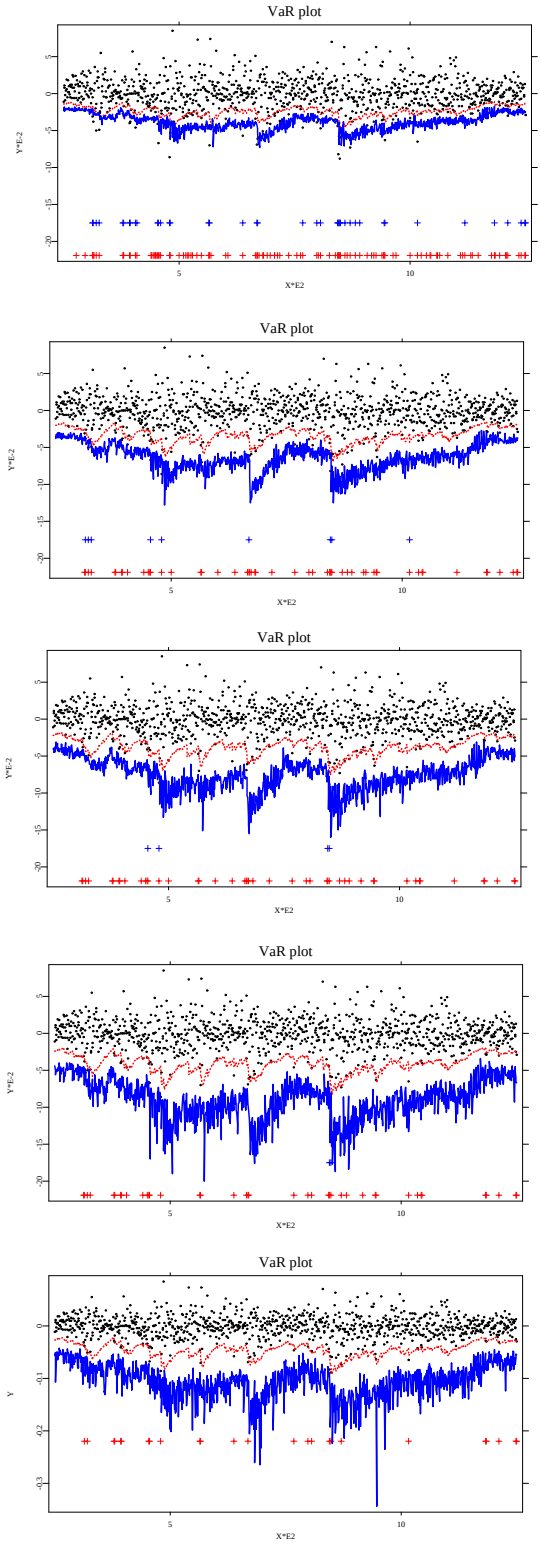

Fig. 5.6: VaR time plots of the exchange rate portfolio with the weight $b = (1, 2)^\top$.

ICAfxVaRplot.xpl

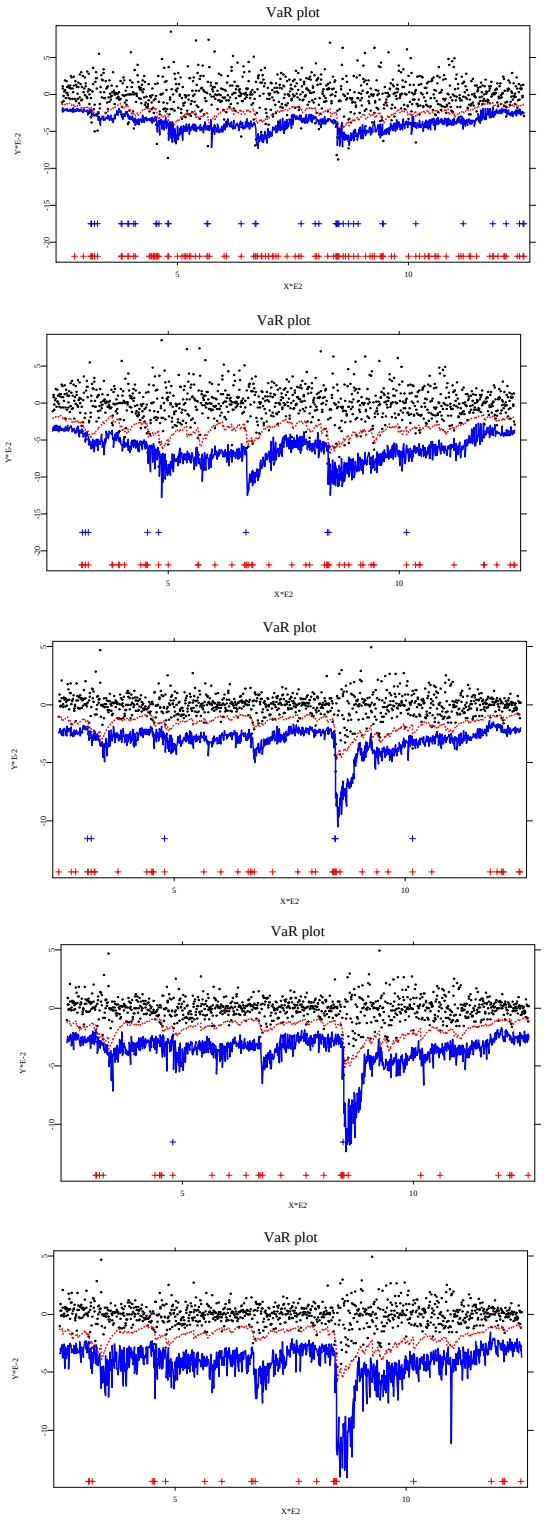

Fig. 5.7: VaR time plots of the exchange rate portfolio with the weight $b = (-1, 2)^\top$.

ICAfxVaRplot.xpl

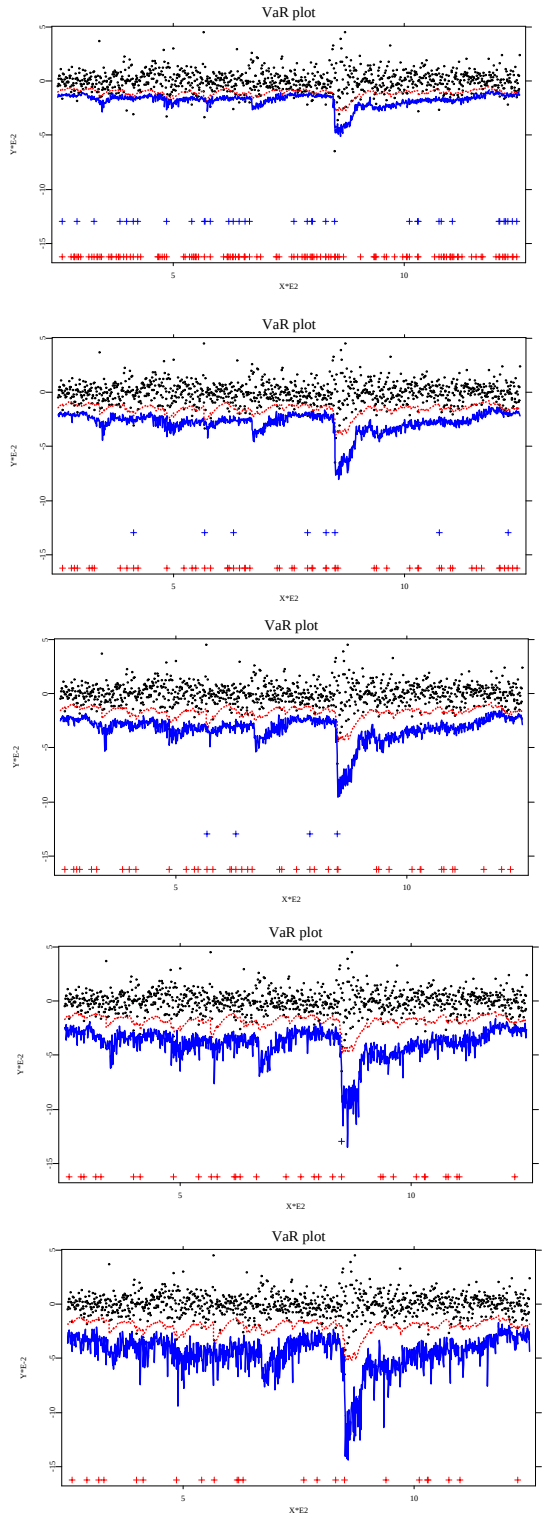

Fig. 5.8: VaR time plots of the exchange rate portfolio with the weight $b = (-2, 1)^\top$.
ICAfxVaRplot.xpl

6. CONCLUSION

In this master thesis, we proposed a multivariate risk management model. The study is mainly based on the independent component analysis (ICA) and a heavy-tailed distribution assumption. The ICA avoids the estimation of the dependence structure of a high-dimensional data and further its joint density in a multivariate case. The linear transformation is found by ICA, which can help to find independent components (ICs). The joint density is the product of the marginal ones and the covariance is a diagonal matrix. In this sense, the multivariate analysis is simply decomposed to the volatility and the marginal density estimations of each IC. The GH distribution is preferable since it mimics the empirical densities of the ICs accurately. Furthermore, the adaptive methodology is used to estimate the dynamic volatility processes. The implementation of the three methods performs outstanding in the VaR study based on a 2-dimensional exchange rate portfolio. Our study is summarized as follows.

- The adaptive volatility estimation methodology by Mercurio and Spokoiny (2003) is applicable for a generalized hyperbolic distribute process. The threshold parameter used to specify the time homogeneity interval can be estimated in a nonparametric way.

- The distribution of the devolatilized returns from the adaptive volatility estimation is found to be leptokurtic. It can be perfectly modelled by the HYP and NIG distributions, subclasses of the generalized hyperbolic distribution.

- The proposed approach can be easily applied to calculate and forecast risk measures such as value at risk and expected shortfall. On the basis of the exchange rate portfolio with several trading strategies. It shows that the proposed approach performs much better than the widely used RiskMetrics model.

BIBLIOGRAPHY

Atkinson, A. C. (1982). The simulation of generalized inverse gaussian and hyperbolic random variables, *SIAM Journal on Scientiffic Computing* **3**: 502–515.

Back, A. and Weigend, A. (1998). A first application of independent component analysis to extracting structure from stock returns, *International Journal of Neural Systems* **8**: 473–484.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size, *Proceedings of the Royal Society of London* **A 353**: 401–419.

Barndorff-Nielsen, O. (1997). Normal inverse gaussian distributions and stochastic volatility modelling, *Scandinavian Journal of Statistics* **24**: 1–13.

Barndorff-Nielsen, O. and Blæsild, P. (1981). Hyperbolic distribution and ramifications: Contributions to theory and applications, *in* C. Taillie, P. Patil and A. Baldessari (eds), *Statistical Distributions in Scientific Work*, Vol. 4, D. Reidel, pp. 19–44.

Bibby, B. M. and Sørensen, M. (2001). *Hyperbolic Processes in Finance, Technical Report 88*, University of Aarhus, Aarhus School of Business.

Bollerslev, T. (1995). Generalied autoregressive conditional heteroskedasticity, *in* R. Engle (ed.), *ARCH, selected readings*, Oxford University Press, pp. 42–60.

Cardoso, J. and Souloumiac, A. (1993). Blind beamforming for non gaussian signals, *IEEE-Proceedings-F* **140**: 362–370.

Chen, Y., Härdle, W. and Jeong, S. (2005). Nonparametric risk management with generalized hyperbolic distributions, *SFB, discussion paper*.

Christoffersen, P. F. (1998). Evaluating interval forecast, *International Economic Review* **39**: 841–862.

Comon, P. (1994). Independent component analysis - a new concept?, *Signal Processing* **36**: 287–314.

Cont, R. and Tankov, P. (2004). *Financial Modelling With Jump Processes*, Chapman & Hall/CRC.

Cover, T. and Thomas, J. (1991). *Elements of Information Theory*, John Wiley & Sons.

Duann, J., Jung, T., Kuo, W., Yeh, t., Makeig, S., Hsieh, J. and Sejnowski, T. (2002). Single-trial variability in event-related bold signals, *NeuroImage* **15**: 823–835.

Eberlein, E. and Keller, U. (1995). Hyperbolic distributions in finance, *Bernoulli* **1**: 281–299.

Eberlein, E., Kallsen, J. and Kristen, J. (2003). Risk management based on stochastic volatility, *Journal of Risk* **5**: 19–44.

Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls, *in* M. Dempster (ed.), *Risk Management: Value at Risk and Beyond*, Cambridge University Press.

Engle, R. F. (1995). *Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, ARCH*, Oxford University Press.

Flury, B. (1998). *Common Principal Components and Related Multivariate Models*, John Wiley & Sons, Inc.

Franke, J., Härdle, W. and Hafner, C. (2004). *Statistics of Financial Markets*, Springer-Verlag Berlin Heidelberg New York.

Härdle, W. and Simar, L. (2003). *Applied Multivariate Statistical Analysis*, Springer-Verlag Berlin Heidelberg New York.

Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and Semiparametric Models*, Springer Verlag.

Harvey, A., Ruiz, E. and Shephard, N. (1995). Multivariate stochastic variance models, *in* R. Engle (ed.), *ARCH, selected readings*, Oxford University Press, pp. 256–276.

Hyvärinen, A. (1998). *New Approximations of Differential Entropy for Independent Component Analysis and Projection Pursuit*, MIT Press, pp. 273–279.

Hyvärinen, A. (1999). Gaussian moments for noisy independent component analysis, *IEEE Signal Processing Letters* **6**: 145–147.

Hyvärinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis, *Neural Computation* **9**: 1483–1492.

Hyvärinen, A. and Oja, E. (1999). Independent component analysis: Algorithms and applications, *Neural Networks* **13**: 411–430.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis*, John Wiley & Sons, Inc.

Jaschke, S. and Jiang, Y. (2002). Approximating value at risk in conditional gaussian models, *in* W. Härdle, T. Kleinow and G. Stahl (eds), *Applied Quantitative Finance*, Springer Verlag.

Jorion, P. (2001). *Value at Risk*, McGraw-Hill.

Mercurio, D. and Spokoiny, V. (2003). Statistical inference for time inhomogeneous volatility models, *submitted*.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. (1992). *Numerical Recipes in C*, Cambridge University Press.

Ristaniemi, T., Raju, K. and Karhunen, J. (2002). Jammer mitigation in ds-cdma using independent component analysis, *In Proc. of the IEEE Int. Conf. on Communications*.

Taleb, N. (2001). *Fooled by Randomness: The Hidden Role of Chance in the Markets and Life*, Texere Publishing.